Exploring the Viability of Meta-Models for Fast and Accurate Prediction of the Behavior of Variable Shape Mould Systems

Christensen, Esben Toke

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

EXPLORING THE VIABILITY OF META-MODELS FOR FAST AND ACCURATE PREDICTION OF THE BEHAVIOR OF VARIABLE SHAPE MOULD SYSTEMS

BY
ESBEN TOKE CHRISTENSEN

DISSERTATION SUBMITTED 2019

AALBORG UNIVERSITY
DENMARK

Exploring the Viability of Meta-Models for Fast and Accurate Prediction of the Behavior of Variable Shape Mould Systems

Ph.D. Thesis

by

Esben Toke Christensen

Department of Materials and Production, Aalborg University, Fibigerstræde 16, DK-9220 Aalborg Øst, Denmark

Dissertation submitted May 26th, 2019

Dissertation submitted: May 26th, 2019

PhD supervisor: Assoc. Prof. Esben Lindgaard
Aalborg University

Assistant PhD supervisor: Prof. Erik Lund
Aalborg University

PhD committee: Professor John Rasmussen (chairman)
Aalborg University

Professor Vassili Toropov
Queen Mary University of London

Associate Professor Jan Becker Høgsberg
Technical University of Denmark

PhD Series: Faculty of Engineering and Science, Aalborg University

Department: Department of Materials and Production

ISSN (online): 2446-1636
ISBN (online): 978-87-7210-442-3

Published by:
Aalborg University Press
Langagervej 2
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2019

Preface

This thesis has been submitted to the Faculty of Engineering and Science at Aalborg University in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Mechanical Engineering. The thesis is written as a collection of papers together with an extended summary, in accordance with the recommendation of the Doctoral School of Engineering and Science.

The Ph.D. project was part of a joint research project between Aalborg University and the Aalborg-based company Adapa. This research project was funded in part by The Danish National Advanced Technology Foundation (Grant No. 48-2012-3). I am grateful for this support.

I want to extend my greatest gratitude towards Erik Lund and Esben Lindgaard, my supervisors during this project. They have been a great source of inspiration, guidance, and support, especially through the hard times. I also want to thank Alexander I. J. Forrester for welcoming and guiding me during my stay at The University of Southampton.

I further want to express my gratitude to my colleague and dear friend through many years, Jens Ammitzbøll Glud. He has been a constant source of support, help and good company in our shared office at Aalborg University, and during our stay in Southampton.

Last but not least, I want to thank my dear wife Lykke for always being there for me, and for enduring my time spent working, writing, and worrying.

Esben Toke Christensen
Aalborg, May 26th 2019

Abstract

Mass customization in production is increasing in demand in almost every industry. In order to meet this demand there is a need for new tooling. It is simply too expensive to create purpose-made tooling for producing low counts of a particular part.

Variable shape moulds represent one solution for addressing this issue. A variable shape mould is a motorized mould that can change shape quickly, enabling production of many different specimens on the same piece of equipment. The focus of this Ph.D. thesis is the mechanical characterization of such a system, and the development of numerical tools for quickly and accurately determining the actuator positions required for producing a given shape on this system. The latter is done through the development of surrogate models for predicting the behaviour of the system in conjunction with an optimization method for finding the actuator positions.

The first part of the thesis is provided as a general introduction to the field of study and to provide context and guidance for reading the included papers. The introduction starts with an overview of the field of variable shape moulds, and a more detailed presentation of the specific system studied in the project. Following this the challenges and research questions of the project are formulated. An overview of the work done with respect to the mechanical characterization of the system is then presented. A general introduction to the field of surrogate modelling is given, followed by a brief presentation of the work done with respect to optimization. Finally, an introduction to the three journal papers that form the basis of this thesis is given, followed by a summary of the scientific contributions of the project and suggestions for future work.

The papers that constitute the main part of the thesis are found in part 2. In paper A four surrogate models are developed. These are difference methods where surrogate models based on Kriging and proper orthogonal decomposition are used to correct a numerically efficient but inaccurate spline based model. Two of the models employ a novel patch-wise modelling scheme. The models are trained and benchmarked on data obtained from a simple finite element model representing a variable shape mould system. Paper B continues the work in paper A by validating the developed models on measurement data from two variable shape mould prototypes obtained using digital image correlation. Paper C introduces a novel method for generating constrained orthogonal-maximin nested Latin hypercube sample designs.

Resumé

Industrien efterspørger i højere og højere grad massetilpasning i produktionen. For at imødekomme denne efterspørgsel er der behov for nyt produktionsudstyr. Det er ganske enkelt for dyrt at fremstille formålsspecifikt udstyr til fremstilling af komponenter ved lave styktal.

Variable forme udgør én løsning på dette problem. En variabel form er en motoriseret form der kan skifte facon på kort tid, hvilket muliggør fremstilling af mange forskellige emner på det samme stykke udstyr. Denne afhandling fokuserer på mekanisk karakterisering af et sådant system, samt udvikling af numeriske værktøjer til hurtig og præcis bestemmelse af de nødvendige motorpositioner til fremstilling af et givet emne på formen. Det sidstnævnte er opnået ved brug af optimering sammen med surrogat-modeller udviklet til formålet.

Den første del af denne afhandling fungerer som en generel introduktion til emnet, og tilsigter at etablere kontekst for de inkluderede artikler. Introduktionen starter med et overblik over variable forme generelt, og det specifikke system der arbejdes med i afhandlingen præsenteres i flere detaljer. Derefter formuleres projektets udfordringer og forskningsspørgsmål. Der gives herefter en introduktion til det arbejde der er udført med henblik på mekanisk karakterisering af systemet der arbejdes med. En generel introduktion til emnet surrogatmodellering gives derefter, efterfulgt af en præsentation af det udførte arbejde indenfor optimering. Til slut præsenteres de tre vedlagte artikler, de videnskabelige bidrag opsummeres og der gives forslag til videre arbejde indenfor emnet.

De tre artikler der udgør hoveddelen af denne afhandling kan findes i del 2. I artikel A udvikles fire surrogatmodeller. Disse er differensmetoder hvor surrogatmodeller baseret på Kriging of proper orthogonal decomposition benyttes til at korrigere en numerisk effektiv men unøjagtig spline-baseret model. To af modellerne benytter sig af en ny områdevis modelstruktur. Modellerne er trænet og testet på data fra en simpel finite element model der repræsenterer en variabel form. Artikel B fortsætter arbejdet i artikel A ved at validere de udviklede modeller på data fra to prototyper af variable forme, målt ved brug af digital image correlation. I artikel C introduceres en ny metode til at generere constrained orthogonal-maximin nested Latin hypercubes.

Thesis Details

This thesis is submitted as a collection of papers, and consists of an introduction to the area of research and the context of the project, and three papers for publication in refereed scientific journals. Currently, two of these papers are published, and the third is to be submitted.

Thesis Title:	Exploring the Viability of Meta-Models for Fast and Accurate Prediction of the Behavior of Variable Shape Mould Systems
Ph.D. Student:	Esben Toke Christensen
Ph.D. Supervisors:	Esben Lindgaard, Assoc. Prof., Ph.D., M.Sc. Department of Materials and Production Aalborg University, Denmark
	Erik Lund, Prof., Ph.D., M.Sc. Department of Materials and Production Aalborg University, Denmark

Publications in Refereed Journals

A) Christensen, E. T., Forrester, A. I. J., Lund, E., Lindgaard, E. 2018. Developing Metamodels for Fast and Accurate Prediction of the Draping of Physical Surfaces, *Journal of Computing and Information Science in Engineering*, **18**(2), 021003.

B) Christensen, E. T., Lund, E., Lindgaard, E. 2018. Experimental Validation of Surrogate Models for Predicting the Draping of Physical Interpolating Surfaces, *Journal of Mechanical Design*, **140**(1), 011401.

C) Christensen, E. T., Lund, E., Lindgaard, E. 2018. Constrained Orthogonal-Maximin Nested Latin Hypercube Sampling, *to be submitted*.

Publications in Proceedings with Review

D) Christensen, E. T., Glud, J. A., Lindgaard, E., Bak, B. L. V. 2013. Analysing the strength of wrinkle defects in glass-epoxy laminates, *Proceedings of the 26th Nordic Seminar on Computational Mechanics.*

E) Glud, J. A., Christensen, E. T., Lindgaard, E., Bak, B. L. V. 2013. Implementation of a state-of-the-art cohesive zone element for ANSYS Mechanical, *Proceedings of the 26th Nordic Seminar on Computational Mechanics.*

F) Christensen, E. T., Forrester, A. I. J., Lund, E., Lindgaard, E. 2015. Using Meta-Models as Fast and Accurate Predictors of a Reconfigurable Mould System, *Proceedings of the ASME 2015 International Mechanical Engineering Congress & Exposition.*[1]

This thesis has been submitted for assessment in partial fulfilment of the requirements of the Ph.D. degree. The thesis is based on the submitted or published papers listed above. Parts of the papers have been used directly or indirectly in the extended summary of the thesis.

[1]The conference paper was accepted, but was pulled back prior to publication, due to concerns w.r.t. protection of Adapa's intellectual property.

Contents

Part I

Introduction

Chapter 1

Introduction

In this chapter an introduction to the Ph.D. project is given, and the field of research and the research objectives are established. A general introduction to the field of research is given in order to present the challenges that lead to the research objectives of the project. Following this a brief overview of the work that has been carried out during the project is presented. The purpose of this chapter is therefore to establish the context of this Ph.D. thesis, and to give the reader a sense of the chronology of the project in a broader sense.

1.1 Background

This Ph.D. project has been carried out as a joint research project between the Danish National Technology Foundation (DNATF), Aalborg University (AAU) and the Aalborg-based company Adapa Aps. Adapa's field of expertise is development and production of variable shape mould systems, which was the focus area in this research project. The overall goal of the project was the further development of Adapa's variable shape mould technology. The role of Aalborg University within this project was analysis of the existing technology, and to develop efficient methods for predicting the behaviour of such mould systems, with the purpose of using these predictions for improving production tolerances of the specimen made using the mould system. This Ph.D. project was initiated within this topic.

The outcome of the Ph.D. project is threefold. One part is the transfer of the generated knowledge to Adapa. Another is knowledge dissemination through conference participation and teaching. The final part is the academic outcome, which is documented in three research papers.

1.2 Variable Shape Moulds

In a market that, to a larger and larger extent, demands complex and unique products, the need to establish fast, flexible, and cost-effective production systems for mass customization is evident. In many industries dealing with large surfaces, e.g. building facades, ship hulls, and wind turbine blades, 3D double-curved panels are of significant importance. Currently, to produce such panels, a separate mould must be created for each unique panel shape. Such moulds are often created by advanced 3D milling processes followed by post-milling surface finishing, in order to reach the required shape and surface quality. This requires a great amount of manual labour and large production facilities. For low production quantities or rapid prototyping for research and development this will result in high cost, both financially and environmentally. Due to this significant research has been carried out in the field of variable shape tooling over the years.

Broadly formulated variable shape tooling is tooling whose shape can be changed within a short time frame, according to need. Many such systems have been proposed in the literature, but in the following a more specific focus is used; variable shape moulds with the purpose of producing a surface, on which to cast or form material. Tooling of this type will be referred to as variable shape moulds. Variable shape moulds can roughly be separated into two categories: moulds with close-packed pins, and moulds with distributed pins. For both types an array of actuated pins constitutes the basic concept of the mould, see Figure 1.1. Each pin serves as a discrete point in the plane (x, y) in which the height of the mould surface (z) can be prescribed by moving the pin up and down. For moulds with distributed pins, an interpolating membrane that provides the mould shape between pins is necessary. For moulds with close-packed pins this may not be necessary, but can be used to suppress the discrete nature of the mould. The differences between different ideas in the field mainly consist of the choices of pin actuation, pin fixation, and surface interpolation. In the following a short overview of literature covering each system type is given.

1.2.1 Systems with Close-Packed Pins

A research project at Massachusetts Institute of Technology (MIT) running from 1980 into the late 1990's led by David Hardt dealt with the development of a variable shape mould system primarily for sheet metal forming [1]. In cooperation with Daniel Walczyk, experience obtained through this project were summarized in [2] where guidelines for the design of such systems are given. Especially pin shape and size, method of pin fixation and pin actuation were considered in detail. Square pins with hemispherically rounded tips were recommended for the ease of assembly in the die, cost of manu-

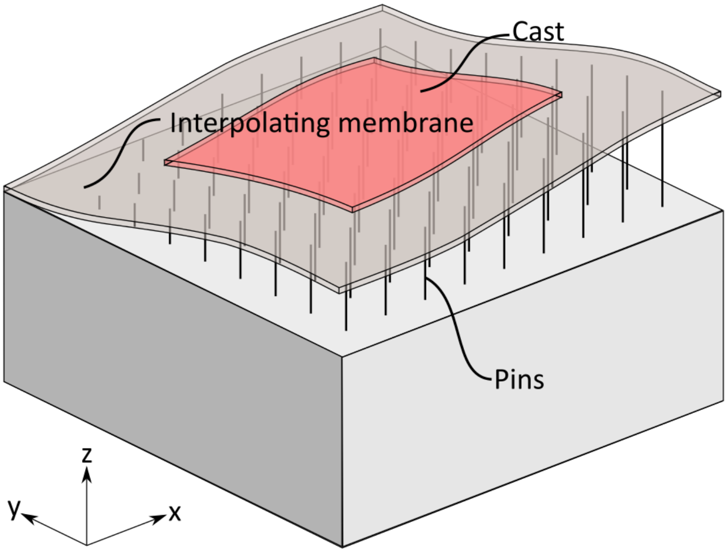

Fig. 1.1: The working principle of the considered type of variable shape mould.

facturing, and their well-defined behaviour when side clamped. Pin size of course has a direct influence on the level of detail that can be supplied by the mould and should be chosen according to needs. In [2] it is argued that, due to tolerance build-up, the total mould size and pin sizes are to some degree dependent. For small pins actuation is suggested to be carried out pin by pin or row by row with a movable profiling mechanism. For larger pins individual pin actuation using either a motor-driven lead-screw or hydraulics is suggested. With individual pin actuation, pin fixation can be handled individually. For row by row pin setting, each pin is in principle free to move. A side clamping method is suggested in this case, where a large force is applied to the pin matrix, locking the pins in place by pin to pin friction. In order to prevent pins from affecting each other during pin movement, sheet metal dividers, which are connected to the mould frame and running perpendicular to the direction of clamping, are suggested. These have the purpose of separating individual rows of pins. It is argued that these dividers also help transferring the reaction force of the forming load to the outer frame. The use of a sheet of Elvax[1] with a thickness equal to the pin size was recommended as an interpolator in order to effectively suppress dimpling of the sheet metal parts. During the research project a few patents were also obtained. The patents [3,4] describe the principle of a computer controlled die with individually actuated pins and a pin actuation module respectively. The

[1] A Dupont product: Ethylene Vinyl Acetate Copolymer.

research project conducted at MIT ultimately led to the commercially available Reconfigurable Tooling for Flexible Fabrication (RTFF) die manufactured by Cyril Bath, who still offers a similar product for sheet metal stretch forming. In the early efforts of the research project the use of a reconfigurable die for sheet metal forming together with a feedback control scheme, for minimizing the part shape error due to material springback and variability in material properties, was demonstrated [5]. This was further developed later in the project [6]. Although the above is judged to still largely represent the state-of-the-art in close-packed adjustable moulds a few later contributions are worth mentioning. In [7] the use of a discrete die for incremental diaphragm-forming of double-curved composite panels is demonstrated. The term 'incremental' in this case means that the shape of the adjustable mould is changed during the process in order to improve shapeability of the compound. In [8] a software tool for determining the required pin heights in discrete dies from CAD-geometry of a given part was developed. This work covered both conventional pin-type dies as well as a commercial die by Surface Generation Ltd [2]. In [9] the possibilities of using an arrangement of six close-packed adjustable dies (one for each side of a cube) in order to create a three-dimensional adjustable mould cavity for injection moulding is discussed.

1.2.2 Systems with Distributed Pins

Less emphasis has historically been put on variable shape moulds with distributed pins in the literature. This is evident from the fact that reviews such as [1] only mentions two publications on the matter. However, in the recent years, substantial work has been carried out - especially in the field of architecture. In [10] a method for annealing freely curved glass panels is demonstrated. The method has later been patented [11]. The mould consists of a number of CNC cut support plates which define the mould shape at various sections along the length of the mould. The interpolator consists of a series of steel rods with stiffnesses chosen according to glass thickness, intended glass panel curvature and support plate spacing. A system that allows the mould to remain plane until the plane glass panel has been heated to its forming temperature is also described. Although the developed mould relies on CNC cut support plates for providing the mould shape and thus is not strictly adjustable, replacement of these support plates with a grid of linear actuators is suggested for future research in the project. Belis et al. later used a similar technique, using a stainless steel mesh interpolator [12]. In a continuation of the work in [10], the working principle is adapted to the process of casting concrete panels in a research project at the Delft University of Technology. This research project centers around a Ph.D. project on the subject by Roel

[2] Surface Generation no longer advertises this tool on their web-page.

Schipper. In this work an array of linear actuators is used to shape the mould and several different types of interpolators have been used [13][3].

1.2.3 Commercial Applications

Through the cooperative research of MIT, the Northrop Grumman Corporation and Cyril Bath, the previously mentioned commercially available RTFF was developed [1] based on the patents [14], [4]. The RTFF die used a close-packed array of leadscrew actuated square pins and utilized an interpolating urethane membrane. The first tool that was produced was sold to the Warner-Robbins Air Logistics Center in 2002 for stretch forming purposes [1]. As previously mentioned a similar product is still offered by Cyril Bath for sheet metal stretch forming. The Pinbed Wizard[4] by FD Technologies offers an array of distributed linear actuators, that can be adjusted in order to collectively represent the shape of the surface to be cast. A different interpolator is used depending on the requirements of the product to be manufactured, and is thus not part of the product. It is not apparent whether the machine is being sold, but the company offers consultancy in production using adaptive moulds and in production of double-curved geometries in general. North Sails uses adjustable moulds for production of their 3DL and 3Di sails[5]. The moulds are so large that an entire fiber-reinforced sail can be laid up and cured in one step. These moulds are also comprised of an array of linear actuators, but here the interpolator is an integral part of the mould, and seems to be mechanized rather than a simple elastic surface. Surface Generation Ltd. offers their patented Reconfigurable Pin Tooling system [14]. The system is reminiscent of that of the RTFF die with a close-packed array of square pins, which can be adjusted individually. However in this system the top halves of the pins are replaceable and are made from a machinable material. Upon adjusting the mould to the wanted shape but with positive tolerances, the surface can be machined to meet the exact required shape [1], [8]. In this way the advantages of a CNC milled mould can be obtained with less material waste.

1.3 The AdaptiveMould

The system presented in the following, which will be referred to as *the AdaptiveMould*, is the mould prototype as it appeared at the beginning of the Ph.D. project. Since then, the system has undergone numerous design iterations, greatly improving the performance and robustness of the system.

[3] See also `http://homepage.tudelft.nl/6w3a0/FlexibleMouldProject.htm`
[4] See `http://fdtechnologies.eu`
[5] See `https://northsails.com/`

The AdaptiveMould by Adapa was the technology to be matured in the research project that this Ph.D. project was part of. The AdaptiveMould is shown in Figure 1.2 and consists of three main parts:

1. A bed of actuators.
2. An interface between actuators and the interpolator.
3. An interpolating membrane.

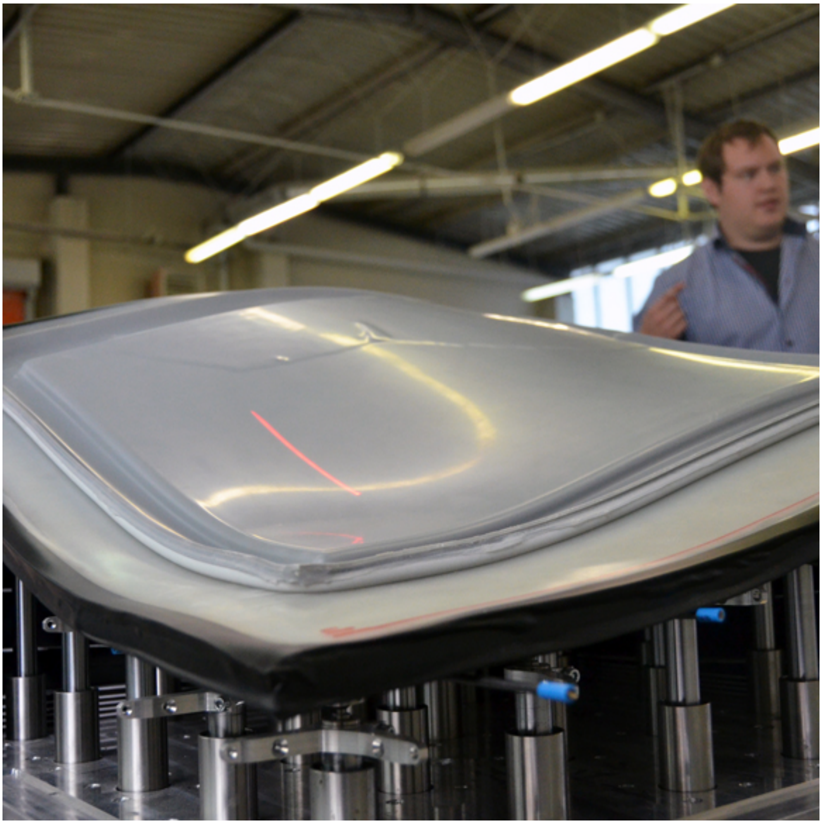

Fig. 1.2: The AdaptiveMould in action.

1.3.1 Bed of Actuators

The AdaptiveMould is a distributed pin variable shape mould with a 9 by 9 square-grid array of stepper-ball-screw actuators. As described in section 1.2.2, the term ’ditributed pin’ refers to the fact that the actuators are placed some distance apart. In this case the actuators are placed periodically, 250mm

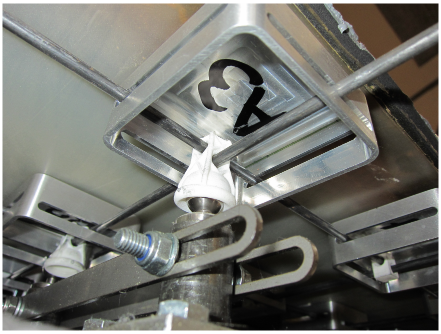

Fig. 1.3: A bottom up view of the actuator to membrane interface.

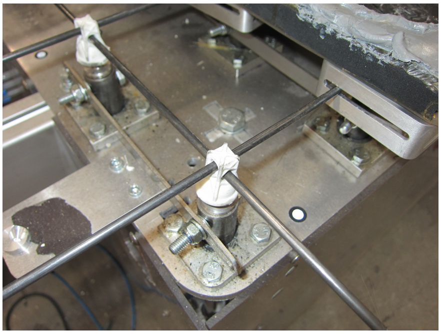

Fig. 1.4: At the edge of the mould, outside of the membrane, a top-down view of part of the interface is possible.

apart. The large distance between actuators allows for using fewer actuators, but limits the mould degrees of freedom and places a high demand on the quality of the interpolating membrane to ensure good surface tolerances. The choice of stepper-ball-screw actuators ensures that the actuator displacements can be set with high accuracy and resolution.

1.3.2 The Actuator-to-Membrane Interface

In Figures 1.3 and 1.4 the actuator to membrane interface is shown. The white plastic part in the center of each image serves as the intersection between rows of spring steel bars following each row of actuators. These plastic parts are connected to the actuators in a spherical joint, allowing the spring bars to act as splines interpolating over each row of actuators. In Figure 1.3 a slitted metal box is seen. This box is fixed to the interpolating membrane, and the spring steel bars are run through the slits. This connection allows translation along the tangents of the spring steel bars.

The combination of the spherical and translational joints ensures that the movement of the interpolating membrane relative to the actuator is only restricted along the membrane normal.

1.3.3 The Interpolating Membrane

The interpolating membrane is a complicated, layered structure consisting of a large number of parts, as described in the patent application [15]. The innermost structure consists of a large number of rhombus shaped plastic parts, as shown in Figure 1.6 (Figure 1.5, markings 10 and 10′). These rhombi are placed in an even number of close-packed layers. The layers are pairwise connected by hinges between the individual rhombi (see Figure 1.6). The

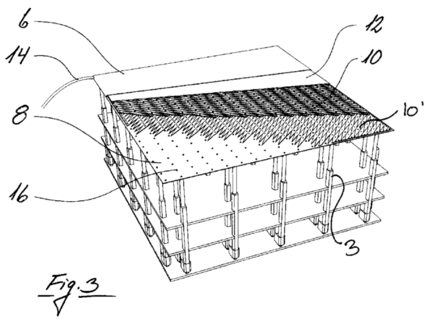

Fig. 1.5: Schematic of the interpolating membrane [15].

Fig. 1.6: Photo of the internal membrane structure.

purpose of the rhombi is to allow the membrane to stretch easily with minor thickness change, while maintaining a good membrane bending stiffness.

On top of the layers of plastic rhombi, a soft neoprene or silicone sheet (Figure 1.5, marking 12) is placed with the purpose of suppressing mould surface dimpling due to the sharp ends of the rhombi or the gaps between them. The whole stack of layers discussed so far is finally hermetically sealed between two layers of silicone (Figure 1.5, markings 6 and 8). When the mould is in use, a vacuum is maintained in the inside of the membrane to keep the internal structure in place.

1.3.4 Advantages and Limitations

The unique membrane structure and properties together with the membrane-actuator interface allowed for creating double curvature mould surfaces of a good surface quality. However, some dimpling due to the membrane internal structure protruding through the top silicone sheet, and general tolerance build-up in the membrane-actuator interface, resulted in low overall tolerances of the complete system. These tolerance problems provided one of the challenges to be considered in the research project.

1.4 Challenges

Throughout the literature there are common themes regarding the specific challenges connected with using variable shape moulds. The most important are:

1. Limited mould surface shape complexity.

   (a) Due to the inherent low number of degrees of freedom.

   (b) Due to mechanical constraints.

2. Limited representation accuracy.
   (a) Due to sub-optimal pin positioning.
   (b) Due to interpolation and dimpling.
   (c) Due to system tolerances and pin positioning accuracy.

Item 1a is a given limitation, since the number of degrees of freedom (DOF) directly corresponds to the number of actuators in a given mould design. Therefore the space of possible mould geometries is typically larger in variable shape moulds of the close-packed pin type, where the number of independent pins is greater. An example concerning item 1b is limitations in the possible mould surface curvature or slope, or simply the range of movement of the individual pins. This challenge is strictly related to the particular mechanical design of a given mould. Items 2a, 2b, and 2c are connected to some extent. Error due to interpolation is furthermore connected to item 1a. Since the mould surface is controlled in a discrete set of locations, an interpolating membrane is used as the mould surface. Unless the interpolative behaviour of this membrane corresponds exactly to the properties of the product to be cast, this mismatch will result in interpolation error. The same goes for the phenomenon of dimpling, where the interaction of the actuation with the interpolating membrane or the structure of the membrane itself results in local disturbances of the mould surface. With respect to item 2c, poor tolerances in the actuator-membrane interface and low pin positioning accuracy can potentially lead to overall poor mould accuracy, since this may result in a difference between the actual and the desired displacement of the interpolating membrane in an actuated location. Finally, depending on the procedure for determining the required pin positions, pin positioning can potentially be sub-optimal due to inaccurate models, measurement error, choice of optimization criterion and algorithm, etc.

The discussed challenges have all been considered in the research project and distributed between the active partners: Aalborg University and Adapa. Items 1a, 1b, 2b, and 2c have been the primary responsibility of Adapa. This is because these challenges are closely connected to the mechanical design of the moulds, which is subject to continuous development by Adapa. Aalborg University has contributed secondarily to these items in the form of ongoing sharing of findings. Item 2a has been the primary responsibility of Aalborg University, and in the following this item will be elaborated and formulated into the initial research objectives of the project.

1.5 Research Questions

Based on the above, a set of research objectives were developed at the outset of the Ph.D. project. The overall objective was to develop and mature the AdaptiveMould technology into the most advanced piece of equipment for production of 3D double-curved panels, and make it appealing to more industries than is the case today.

In order to achieve this, it was clear that a detailed understanding of a state-of-the-art existing variable shape mould system was needed, to realize the root causes of its current limitations and to help improve its mechanical aspects. Such an understanding could be achieved through the development of a detailed finite element model, that could be used for performing parametric studies with the goal of pinpointing key areas for continued improvement.

Even with close to ideal mechanical behaviour, a good control strategy for determining mould actuator displacements is needed to achieve the best possible production tolerances on the mould. Several possibilities were apparent, but the focus for the project is using surrogate models for this purpose. With a sufficiently accurate surrogate model, this model can be used as the source for feedback data. It was therefore hypothesized that an off-line control strategy, as illustrated in Figure 1.7, that uses optimization for adjusting the actuator positions could be used to quickly and efficiently finding a good set of actuator displacements for reproducing a given shape on the mould.

Based on the above, the research questions were established as follows:

1. Can the performance of the AdaptiveMould be improved through the knowledge obtained by an in-depth mechanical characterization of the system?
2. Is it possible to predict the complex behaviour of the flexible membrane using a numerically efficient surrogate model?
3. Given 1. and 2., can improved production tolerances be obtained from the AdaptiveMould using optimization methods for design and adjustment of the moulds?

The presented research questions were largely treated in chronological order. In the following chapters an introduction to the work done within each research question will be given.

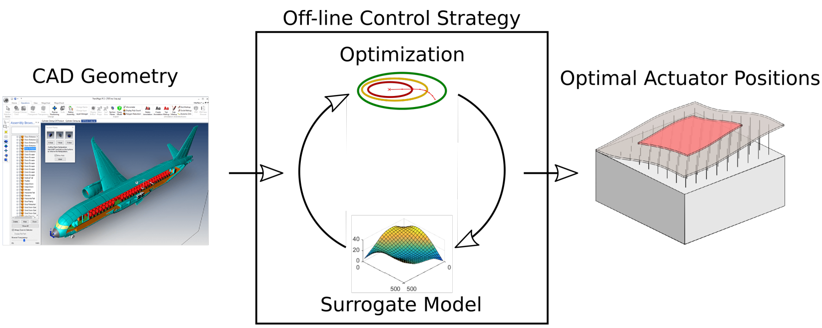

Fig. 1.7: Off-line control strategy to be examined in the research project.

Chapter 2

Mechanical Characterization

The AdaptiveMould described in section 1.3 was the starting point for the present Ph.D. project. In order to address the first research question, a parametrized finite element model of the AdaptiveMould was chosen as a tool, for better understanding the behaviour of the system, and to facilitate its further development. In the following sections the work carried out in this regard will be summarized.

2.1 A Small Membrane

The AdaptiveMould described in section 1.3 was a costly machine which was the constant subject of ongoing development at Adapa's facilities. To be able to examine a cheaper and more stationary subject, a small membrane was manufactured by Adapa, see Figure 2.1. This small membrane would serve as an intermediate target for developing a finite element model against, before expanding the modelling scope to a full size mould.

2.1.1 Geometry

The small membrane was approximately 250mm by 250mm, and consisted of two layers of plastic rhombi between silicone sheets. A double steel frame (visible on Figure 2.1) was manufactured for applying close to fully-clamped boundary conditions on the small membrane, leaving a free membrane area of 210mm by 210mm. This left enough space in the membrane for at least a few non-clamped rhombi. Note that the small membrane cannot be considered a scaled model. This is because the individual parts of the membrane are not scaled down - the size of the rhombi, the silicone sheet thicknesses etc. are the same as those used in the full size membrane. Merely the side length of the membrane are changed.

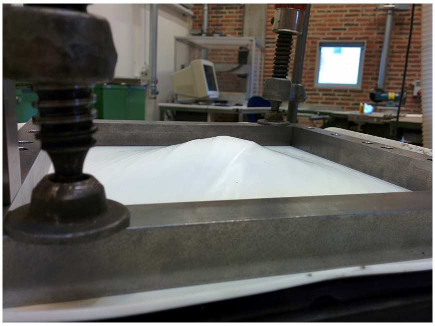

Fig. 2.1: The heavily displaced small membrane.

Fig. 2.2: The center of the small membrane was displaced by a bolt with a ceramic foot.

2.1.2 Constitutive Properties

The needed constitutive properties of the materials used in the small membrane were determined by cutting out small dogbone test specimens, and performing tensile tests to obtain Young's modulus and the Poisson ratios. During use of the interpolating membrane, sliding between the components plays a large role. Therefore the static frictional coefficients for the following material pairs were measured: ABS/ABS (rhombi) and ABS/silicone. Due to the lack of proper equipment for measuring these coefficients, the measurements were performed using a make-shift tilted plane experiment. It was recognized that the accuracy of these measurements would be limited and that ideally, also the dynamic frictional coefficient should be measured. However, it was expected that these properties would need tuning during model development no matter the accuracy, and so the measurements would only serve as a starting point.

2.1.3 Boundary Conditions and Load

The boundary conditions were established as follows: The entire edge of the small membrane was specified as being clamped. This boundary condition was sought physically enforced on the miniature by clamping the edges between two steel frames that were manufactured for the purpose. See Figure 2.1. The center of the miniature membrane was displaced a prescribed displacement of 25mm, using a bolt and a ceramic foot. See Figure 2.2. Note that the applied displacement of 25mm results in a radius of curvature that is very extreme compared to those experienced in use of the AdaptiveMould. An additional load was applied in the form of an internal vacuum of 0.8 bar in the membrane.

Fig. 2.3: Setup for measuring the mould surface shape using digital image correlation.

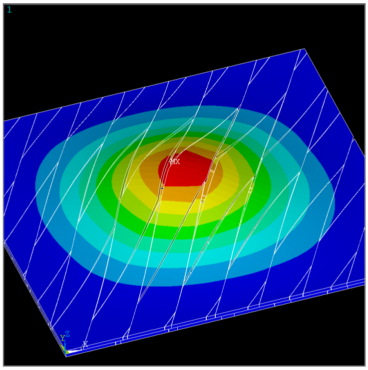

Fig. 2.4: Example of displacement result of the developed finite element model. The top silicone sheet has been hidden to reveal the behavior of the rhombi inside the membrane.

2.1.4 Response Measurement

The resulting shape of the membrane under the conditions described above was measured using stereo digital image correlation. An ARAMIS 4M system was used for this purpose. The speckle pattern was applied directly to the white silicone top sheet of the miniature, using black spray paint and a makeshift nozzle. The digital image correlation measurement setup is shown in Figure 2.3. Using this technique for measuring the deformed shape of the small membrane resulted in a dense 3D surface measurement allowing detailed comparison with the developed finite element model.

2.2 The Developed Finite Element Model

The finite element model was developed in the commercial software, ANSYS[1]. Reproducible model building in ANSYS is done using the built-in scripting language ANSYS parametric design language (APDL). However, due to the complexity needed for modeling the miniature, and with respect to future code reuse when an even larger model of (possibly) a full scale AdaptiveMould would be needed, development directly in APDL was simply not practical.

To overcome this, a model building framework was developed in Matlab. This framework was developed as an object oriented environment that allowed procedurally creating the model, using the powerful tools available in Matlab. This included model geometry, meshing, specification and assign-

[1] `https://www.ansys.com/`

ment of constitutive properties and automatic detection and creation of the needed contact definitions in the model. In Figure 2.4 an example of the displacement result of a converged model is shown, where the top silicone layer is hidden to reveal the internal plastic rhombi. For this model size, the model specific Matlab code used for generating the model totalled approximately 100 lines of code, whereas the generated APDL code was almost 50,000 lines.

Parametric studies were carried out for the developed finite element model, in order to determine the effects of varying material characteristics (stiffnesses, Poisson ratios, frictional coefficients) and the level of applied vacuum. This resulted in a good understanding of the behaviour and important properties of the interpolating membrane. The most important gained understanding was, that a high bending to membrane stiffness ratio is advantageous for the performance of the interpolating membrane. Inuitively this can be understood by thinking of tent cloth stretched over a tent pole. The tent cloth has almost no bending stiffness but a good membrane stiffness. Therefore the cloth will bend excessively close to the pole and little elsewhere, causing the appearance of a very non-smooth ground-pole-ground interpolation. This knowledge was used in the further development of the AdaptiveMould.

Through the study on the small membrane it was realized that in order to reach strong conclusions about the behaviour of the system, a larger membrane was needed. Due to the small distances between load introductions on the small membrane, it proved difficult to discern between local effects from these, and the detailed interpolative behaviour of the membrane. Therefore it was decided to proceed with larger models for further studies.

2.3 Full-Scale Models

A lot of work went into expanding the model to a larger size membrane, with many problems resulting. The primary difficulties concerned problems with getting the larger models to converge, and excessive computational resource requirements. It was soon concluded that the benefits of such a model were outweighed by the resources required for developing it, and a new direction for the project was chosen instead.

Chapter 3

Introduction to Surrogate Modelling

In order to understand choices made from this point on in the project, an introduction to the field of surrogate modelling is required. In the following a general description of what a surrogate model is and what it requires to establish one is presented. Furthermore, a short review of the studied literature in the field is given.

3.1 Introduction

In the studied literature the terms 'surrogate model' and 'meta-model' are used interchangeably. Both terms hint at the way these types of models are often used. This category of models is often used as 'models of models' (meta) or as substitutes (surrogates) for a model, system, measurement etc. In other disciplines several of the methods discussed in the following fall under the terminological umbrella of machine learning.

3.1.1 A Simple Example

Regardless of the terminology, what this refers to is in fact a discipline that many people will know as curve-fitting. In surrogate modelling, the goal is to pick a particular function from some family of functions, that describes the data we are interested in, and therefore allows us to make predictions.

Consider the example shown in Figure 3.1. An unknown, true function (blue line) represents the behaviour of interest. Using a number of noisy samples (blue crosses) measured from the true function, we want to create a model for predicting this function. In this case, the family of functions, from

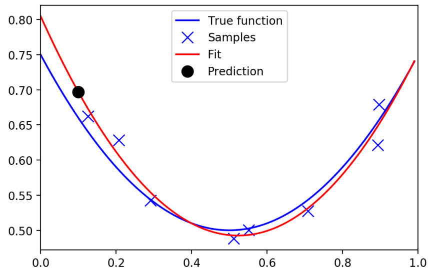

Fig. 3.1: An example curve fit in one dimension.

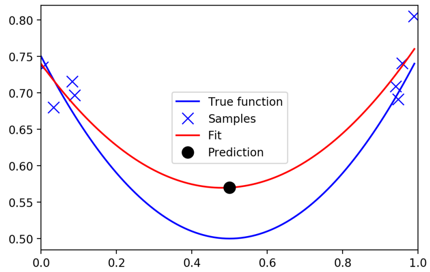

Fig. 3.2: Another curve fit in one dimension. Here the sampling has obviously gone wrong, resulting in a poor model.

which we will select a model, is the family of second degree polynomials. The act of fitting (or training) corresponds to the selection of a particular function from this family that represents the data as well as possible (red line). This newly obtained model can then be used to make predictions of the behaviour of the true function (black circle). In Figure 3.2 the same example is repeated, but with a poor sampling of the true function. It is apparent that proper sampling is important for building a good model.

The presented example demonstrates one of the simplest forms of surrogate modelling in a one dimensional space. Nonetheless, this example illustrates almost all important aspects of surrogate modelling: In order to establish a surrogate model we need data which is sampled properly, we need a family of functions to select a model from, and we need a way of selecting a good function from this family (a training procedure).

While polynomial regression, which was used in the example, typically works very well in low dimensional spaces, more advanced techniques exist that seek to improve modelling performance for higher input dimensionalities and more complex data.

3.1.2 Methods

A large number of model types exist in the literature within the topic of surrogate modelling, and indeed these types of models are used for diverse purposes across many fields [16–20]. In the following, general introductions to a number of the methods that were considered for use during the Ph.D. project are provided.

Polynomial Regression

Polynomial regression is also known as the response surface methodology. This method was previously touched upon as an example. In this method, a polynomial response surface, typically of first or second order, is fitted to observed data using a least squares linear regression, see e.g. [21]. This method has been very successful through time, but has some significant limitations: The method does not work well for highly non-linear, multimodal, or high-dimensional problems, mainly because high order polynomials typically behave erratically and therefore rarely inter- or extrapolate well.

Moving Least Squares

Moving least squares methods solve the problems associated with multimodality and non-linearity, compared to polynomial regression. In this method a polynomial model is fitted locally by weighting the closest (to the point of prediction) data according to some decay function. This method is briefly explained in [22] and is also mentioned as a commonly used method in machine learning in [23]. This method is more computationally demanding than ordinary polynomial regression, because a new fit is performed for every prediction.

Multivariate Adaptive Regression Splines

The method of multivariate adaptive regression splines (MARS) provides for yet another polynomial based method. This method however provides an automatic framework for determining the required complexity of the model, in order to provide a good fit to the data. The resulting model will consist of locally defined spline functions which are stitched together. Each spline function consists of sums of products of univariate spline functions. The details of the method can be seen in [24]. As shown in [25] MARS performs well for dense data and is very computationally efficient. A further advantage is interpretability due to the possibility of performing an analysis of variance (ANOVA) like decomposition of the model.

Artificial Neural Networks

Artificial neural networks (ANN), specifically the feed forward perceptron type, is a very diverse regression model. In a typical setup, such a model consists of stacked layers of ”neurons”. Each neuron should be understood as a simple, non-linear function that accepts a number of inputs, either from the input to the model or from the layer below. ANNs are able to approximate high-dimensional and highly non-linear functions. An introduction to simple ANNs can be found in [26]. In recent years, ANNs have experienced a renaissance with the tremendous successes of recurrent and convolutional neural networks and the advent of deep learning.

Radial Basis Function Networks

Radial basis function (RBF) regression can be considered as a special case of a feed forward perceptron ANN with a single hidden layer and a linear output layer. The difference is primarily the choice of activation function in the ANN (radial basis function whose value depends only on distance from some function center in input space) and the fact that RBF regression is often implemented as a non-parametric method, where more training data implies more complexity in the model. For a non-parametric implementation a RBF is typically centered on every data point in the training set, whereas in a parametric formulation, the number of RBF is a fixed hyper parameter and fitting consists of adapting RBF parameters to fit the data. The latter approach is very similar to a standard feed forward neural network. [22] gives a brief introduction to RBF regression. [27] introduces the extended RBF regression, which improves RBF regression by augmenting the model by a set of univariate non-RBF.

Support Vector Regression

A method, which can be seen as a variant of RBF regression, is support vector regression (SVR). SVR can be interpreted as RBF regression where only the RBFs that are most important for representing the data are included in the model. This viewpoint is illustrated in [22]. It can also be viewed simply as a linear regression in an elevated feature space as done in [28]. [28] gives an easy to read introduction to the method. Training a SVR model corresponds to solving a convex quadratic programming problem which can be done efficiently and with a unique solution. [29] compares SVR with other methods.

Gaussian Processes and Kriging

With Gaussian processes (GP) the training data is assumed to be the outcome of a random process - that is, a single sample from a distribution of functions. By specifying a mean and a covariance function for a GP, a prior is put on this random distribution of functions, specifying prior beliefs about which functions are more probable than others. Therefore function characteristics such as smoothness and variance can be specified through the prior. When training the GP, the possible field of underlying functions is narrowed in, leaving only a smaller amount of possible underlying functions. At prediction time, the mean of the remaining distribution of functions is taken. A very thorough and in-depth treatment of GPs is given in [30]. A special form of GP is Kriging, where a specific and very flexible covariance function is used, and where the GP is typically augmented with a form of linear regression. A good introduction to Kriging is given in [31]. Newer developments include

Blind Kriging, as introduced by [32]. GPs can perform well with sparse data provided a good choice of covariance function. Kriging was ultimately applied in the methods developed and benchmarked in [33, 34], see paper A and paper B on pages 49 and 51, respectively.

Hybrid Methods

A host of hybrid methods exists. Typically, hybrid modelling involves combining two or more of the above mentioned methods, or combining one of the above methods with a physics-based method such as a finite element model. Knowledge based methods and ensembles are worth mentioning. In knowledge based methods, a computationally efficient simplified simulation is used together with a curve fitting method. In this way, knowledge from the simulation can be corrected by the curve fitting method to achieve better accuracy than simulation itself can provide, see e.g. [35, 36]. Ensembles build on the observation, that the methods described previously have different strengths. In an ensemble, surrogate models are built using several methods and, when predicting, a weighted average of the individual predictions is created based on error estimates. This typically provides for more robust methods, see e.g. [37].

Another type of hybrid method involves combining dimensionality reduction methods such as proper orthogonal decomposition [38] with a curve fitting method. This was illustrated for proper orthogonal decomposition and Kriging in [39]. This method was applied in [33, 34], see paper A and paper B on pages 49 and 51, respectively.

3.2 Development Strategy

Regardless of the method, building a surrogate model requires training data. In order to obtain this, a method for sampling the data is needed. As illustrated in the example in section 3.1.1 the choice of sampling method can mean the difference between a good and a bad model. Sampling strategy should therefore be seen as a part of model development. Ultimately a novel sampling strategy was developed during the project. This method was applied in [34], see paper B on page 51, and described in detail in [40], see paper C on page 53.

In addition to selecting a surrogate modelling method, the development of a model will involve the selection of a number of hyper parameters. This could for instance be the choice of kernel in a kernel method, or the number and sizes of the layers in an artificial neural network. In order to choose between two candidate models that are different due to choice of method, sampling strategy, or hyper parameters a performance measure is needed.

This measure is used to evaluate and compare the performances of the two models to facilitate an informed choice. Typically the performance measure will evaluate the quality of predictions at points in the input space that were not part of the training set used to build the models. This is done in order to examine whether the models are able to generalize from the training data. The choice of performance measure has a very direct effect on the development of a model. Therefore this choice could also be seen as part of the model development.

The development of a surrogate model can therefore be seen as the repeated application of a development cycle consisting of selecting a sampling method and sampling training data, selecting a surrogate model and training it, and evaluating the model using the chosen performance measure.

Figure 3.3 illustrates the model development strategy used in the project. As seen, the aforementioned development cycle is at the center of the depicted strategy. Due to considerations regarding the availability and cost of measurement data, this innermost development cycle is iterated twice. First, quick development iterations are performed using a generic model (a simple finite element model) as an interim data source for finding an effective sampling- and modelling scheme. This first iteration is shown in blue line in Figure 3.3. When satisfactory performance is achieved, the second iteration is commenced. In the second iteration, measurement data from a physical system is used as data source. The development cycle will now be much slower due to the higher cost of obtaining the needed data. This second iteration is illustrated in red line in Figure 3.3. Finally, when satisfactory performance is achieved in the second iteration, the resulting surrogate model can be used in a model control scheme for the AdaptiveMould system.

The described development strategy is based on the hypothesis that a model developed for a substitute, can be applied to effectively model the system of interest, provided that the substitute is sufficiently similar to this system.

As described above, the selected development strategy is based on two data sources. In the following section the mentioned data sources are described in more detail and are discussed in the context of data sources in general as well, to highlight the reasoning behind the chosen development strategy.

3.3 Data Sources

A surrogate model must be trained on data that is representative of the behaviour to be predicted. Therefore attention to the quality of the training data must be given. At the same time, data comes at a cost with respect to money, time, and effort. In order to strike a balance between these often opposed

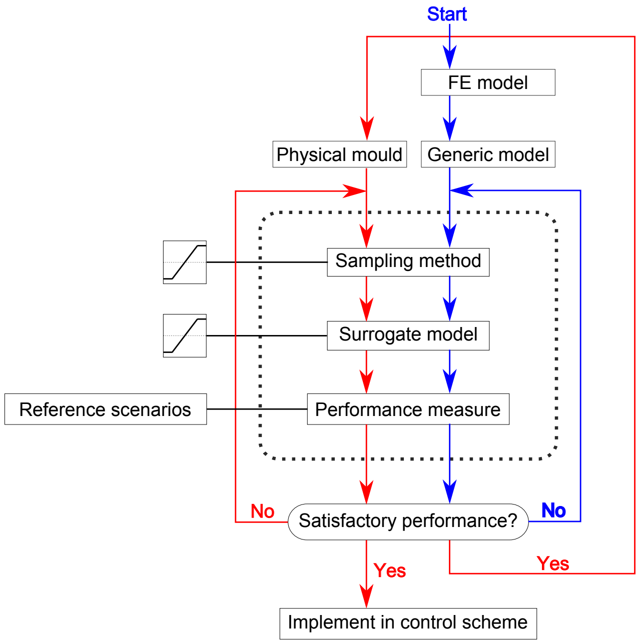

Fig. 3.3: The development process to be carried out for developing an efficient surrogate model.

properties of the data, the following possible data sources for the modelling work were considered.

- Measurements.
- Simulations.
- A substitute.

In the following each of these possibilities are elaborated with respect to data quality and cost.

Measurements

The highest quality data will naturally come from measurements on the system to be modelled. Quality will of course also depend on the measurement method. Stereoscopic Digital image Correlation was the chosen method in this project due to being a fast and accurate method for obtaining full field measurements of a surface shape.

However, measurement data was also the least accessible and a costly data source of the given alternatives. Measurements were not possible during the conducted research stay in Southampton, since the AdaptiveMould prototype had to remain in Aalborg. With respect to cost, the limiting resource was time. Significant effort goes into planning empirical work. Furthermore, due to equipment limitations, measurements could not be automated. Therefore, execution of the measurements required a significant time investment.

Simulations

Simulation data could serve as a substitute for measurement data. The quality of such data would depend on the quality of the used model. Generally speaking, the higher the quality of simulation, the greater the cost (computational time). Even the best model is not perfect. Therefore simulation data should be considered of lower quality than measurement data.

In section 2 the developed finite element models are described. At the time that the surrogate model development strategy was developed, the small finite element model described in section 2.1 could be solved reliably, while it was still unclear whether a larger model was feasible. However, it was clear that so much time would be needed for solving a larger model, that such a simulation would not be a viable data source in a fast development cycle. Due to these considerations, an alternative in the form of a cheap-to-evaluate substitute was chosen.

A Substitute

A generic and simple finite element model was developed to serve as substitute for detailed simulation and to function as a data source for rapid surrogate model development cycles. This model took on the order of seconds to minutes to solve on a laptop and was designed to imitate the overall behaviour of the mould system. Using a model that could be evaluated so quickly facilitated experimenting with sampling methods, since data could be generated on demand in a short time.

The data from the substitute model is inherently of low quality. The substitute was designed with characteristics similar to those of the AdaptiveMould. However these similarities were very general in nature. It was hypothesized that, provided this general similarity, surrogate models developed towards data from the substitute could later be adapted for modelling the AdaptiveMould. A further advantage of using such a model as data source is that it is easier to describe in a research paper, and for others to reproduce.

The generic substitute model is presented in more detail in section 3.3.1 below.

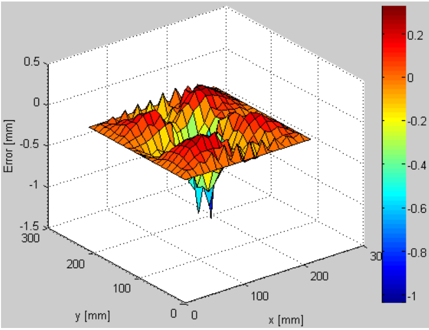

Fig. 3.4: Error in displacement response between the finite element model and the generic model.

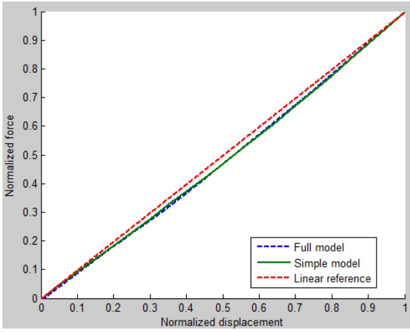

Fig. 3.5: Comparison of the normalized actuator force displacement curves between the finite element model and the generic model.

3.3.1 The Generic Model

The generic substitute model was implemented as a geometrically non-linear finite element simulation in ANSYS. The membrane was represented as a simple plate using first order shear deformation shell elements, and the membrane characteristics were emulated by adjusting the individual entries in the ABDĀ matrix by optimization, with the goal of matching membrane shape and the shape of the actuator force-displacement curve. This parameter tuning was carried out with the goal of fitting the behaviour of the finite element model described in section 2.1. In figures 3.4 and 3.5 the displacement error and force-displacement curves after this matching has been performed are shown. Note that the displacement error close to the point of actuation has been excluded from the objective function, due to the large amount of local effects in this area. Using the material model obtained using this approach, the generic model was established as a finite element model with 25 points of actuation (5 by 5 in a square grid), to represent a small variable shape mould system. The generic model was implemented such that it could be called directly from Matlab.

This model was used in [33, 34], see paper A on page 49 and paper B on page 51. In the former a version with 9 by 9 actuators was used, and in the latter a version that included a simple representation of mechanical freeplay.

Chapter 4

Optimization for Mould Adjustment

In order to illustrate the proposed role of optimization in the project, consider Figure 4.1 (Figure 1.7, page 13 repeated for convenience). In the proposed workflow, a user of the system provides a CAD file specifying the geometry of the part or section to be produced on the AdaptiveMould. This CAD file is then converted to an optimal set of actuator positions for producing this geometry on the AdaptiveMould by off-line control. The term ”off-line” in this context means that the physical AdaptiveMould is not involved in the process. Simply speaking, in a conventional control scheme the needed actuator positions would be found by iteratively measuring the mould shape and readjusting the actuators in proportion to the shape error relative to the CAD geometry. Such a process would be an on-line control strategy.

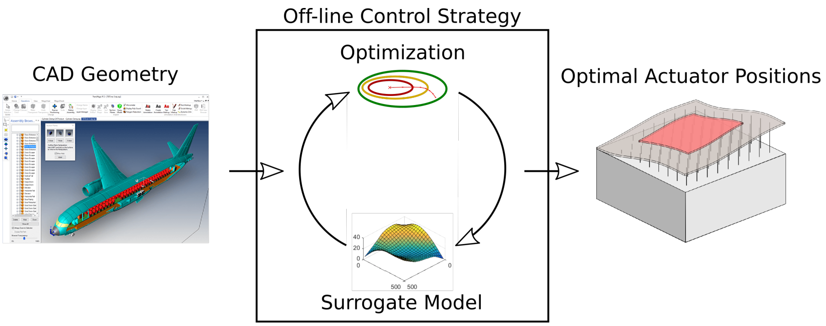

Fig. 4.1: Off-line control strategy to be examined in the research project.

In the proposed off-line control strategy, a surrogate model replaces the physical AdaptiveMould. Finding the optimal actuator positions for a given geometry is then accomplished by formulating an objective function (function of the shape error), constraints (physical limitations and other requirements) and solving the resulting optimization problem.

The papers included in this thesis address research question 2 (see section 1.5 on page 12). Although no research papers have been published regarding research question 3, work was done on this topic during the Ph.D. project. In the following this work is briefly presented.

4.1 A Mould Adjustment Framework

A small Matlab framework for experimentation with optimization for actuator adjustment was developed during the Ph.D. project. The framework allowed for simple assembly of approaches through component-wise selection of model (providing mould behaviour prediction), objective function, and constraints. Furthermore, the framework provided easy to use functions for producing error plots, adjusting objective function weights etc. In section 4.3 an example demonstrating some of the possibilities is given.

4.2 An Efficient Linear Approach

The developed optimization framework allowed for general experimentation. However, at the time it was developed, the surrogate models were not yet ready for use. Therefore an approach closer to the spline based heuristics that Adapa used at the time was developed.

The developed approach uses a predictor based on 1D natural cubic splines that interpolates over the actuators in two directions. While such a function is non-linear with respect to the in-plane coordinates, it is linear with respect to the actuator displacements. Consider a fixed grid, $\mathbf{X}$ in the x-y plane consisting of points $(x_0, y_0), (x_0, y_1), ..., (x_n, y_n)$. The described predictor can be evaluated on such a grid as

$$\bar{w} = \mathbf{D}\bar{h} \tag{4.1}$$

where $\bar{w}$ is a vector representing the predicted shape on $\mathbf{X}$, $\bar{h}$ is a column vector of the actuator heights, and the columns of $\mathbf{D}$ are basis vectors for the predictor on $\mathbf{X}$. It is apparent that, given a left hand side $\hat{w}$ (the wanted shape), Eq. 4.1 is a linear regression problem with the least squares optimal solution [21]

$$\hat{h} = (\mathbf{D}^{\mathrm{T}}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\hat{w} \tag{4.2}$$

Since every derivative of the predictor with respect to the in-plane coordinates will also be linear with respect to the actuator heights, slope and curvature requirements can trivially be included in the problem. If high accuracy at some regions on the mould surface is of particular importance a weight vector, $\hat{m}$, can be used to influence the solution. This is done by solving the weighted least squares regression problem

$$\mathbf{D_M} = \mathbf{MD} \tag{4.3}$$

$$\hat{w}_\mathbf{M} = \mathbf{M}\hat{w} \tag{4.4}$$

$$\hat{h} = (\mathbf{D}_\mathbf{M}^\mathrm{T}\mathbf{D}_\mathbf{M})^{-1}\mathbf{D}_\mathbf{M}^\mathrm{T}\hat{w}_\mathbf{M} \tag{4.5}$$

$\mathbf{M}$ is a diagonal weight matrix where entry $\mathbf{M}_{i,i}$ is the weight associated with the i'th equation.

When introducing system constraints in the form of maximum allowed curvatures or slopes on the mould, an efficient constrained linear least squares optimizer can be used, since these constraints can be formulated linearly as

$$\mathbf{S}\bar{h} \leq \bar{s}_{\mathrm{max}} \tag{4.6}$$

$$-\mathbf{S}\bar{h} \leq \bar{s}_{\mathrm{max}} \tag{4.7}$$

$$\mathbf{C}\bar{h} \leq \bar{c}_{\mathrm{max}} \tag{4.8}$$

$$-\mathbf{C}\bar{h} \leq \bar{c}_{\mathrm{max}} \tag{4.9}$$

Here $\mathbf{S}$ is a matrix whose columns are basis vectors for the spline slopes while the columns of $\mathbf{C}$ are basis vectors for the spline curvatures. $\mathbf{S}$ only describes the slope behaviour at the actuators, since the slope constraints are due to the actuator-to-membrane-interface. $\mathbf{C}$ on the other hand is formulated on $\mathbf{X}$ like $\mathbf{D}$ in Eq. 4.1, since the curvature constraints are due to limitations in the membrane itself. Eqs. 4.7 and 4.9 are needed since the constraints apply to both positive and negative slopes/curvatures.

The described linear approach was developed as an approach that could be used immediately by Adapa, early in the project. Due to the simple formulation of the problem, and the use of off-the-shelf methods, this solution provided for an effective, and easy to implement method with great flexibility for use by Adapa in their products.

4.3 An Example

To demonstrate the capabilities of the developed framework an example is given in the following.

A rectangular panel with a side length of 400mm with a centered 200mm by 200mm cutout is to be produced. The shape of the panel is further deter-

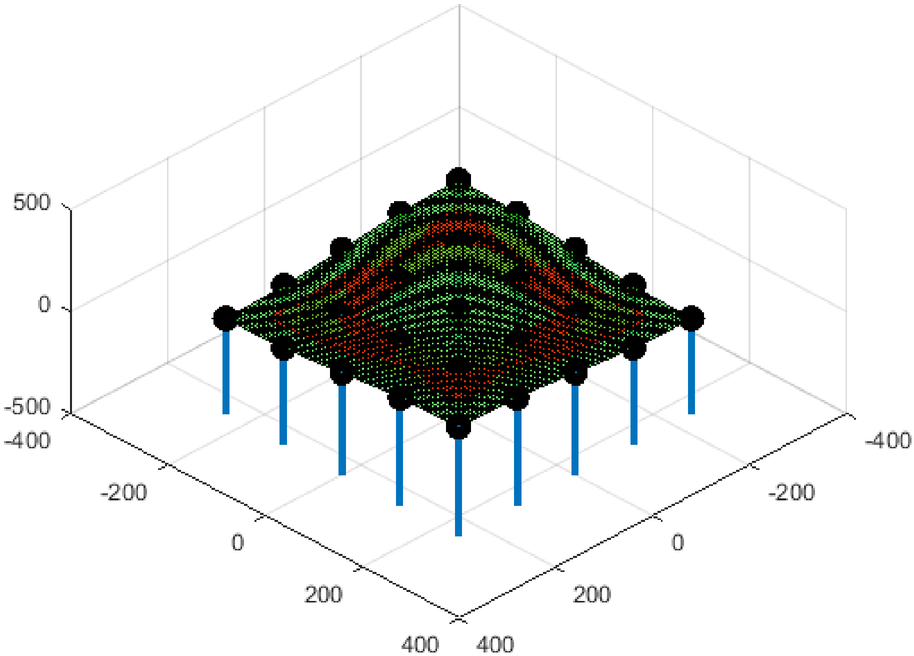

Fig. 4.2: 3D visualization of the mould, as configured to reproduce the defined panel.

mined by

$$z = 32.5 \sin\left(\frac{\pi x}{400}\right) \sin\left(\frac{\pi y}{400}\right) + \left(\frac{x-100}{300}\right)^2 + \left(\frac{y-100}{300}\right)^2 \tag{4.10}$$

that specifies the vertical (z) component. Note that this expression carries no

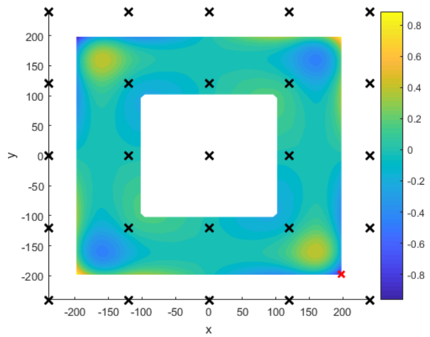

Fig. 4.3: An error plot showing the discrepancy between the predicted mould shape and the shape of the defined panel, for the unconstrained solution.

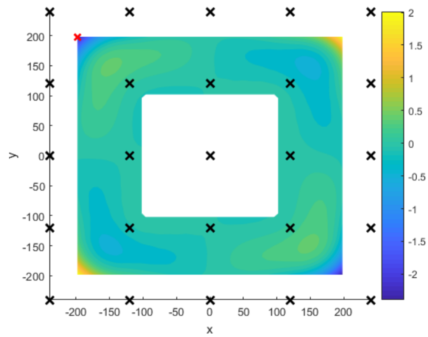

Fig. 4.4: An error plot showing the discrepancy between the predicted mould shape and the shape of the defined panel for the constrained solution.

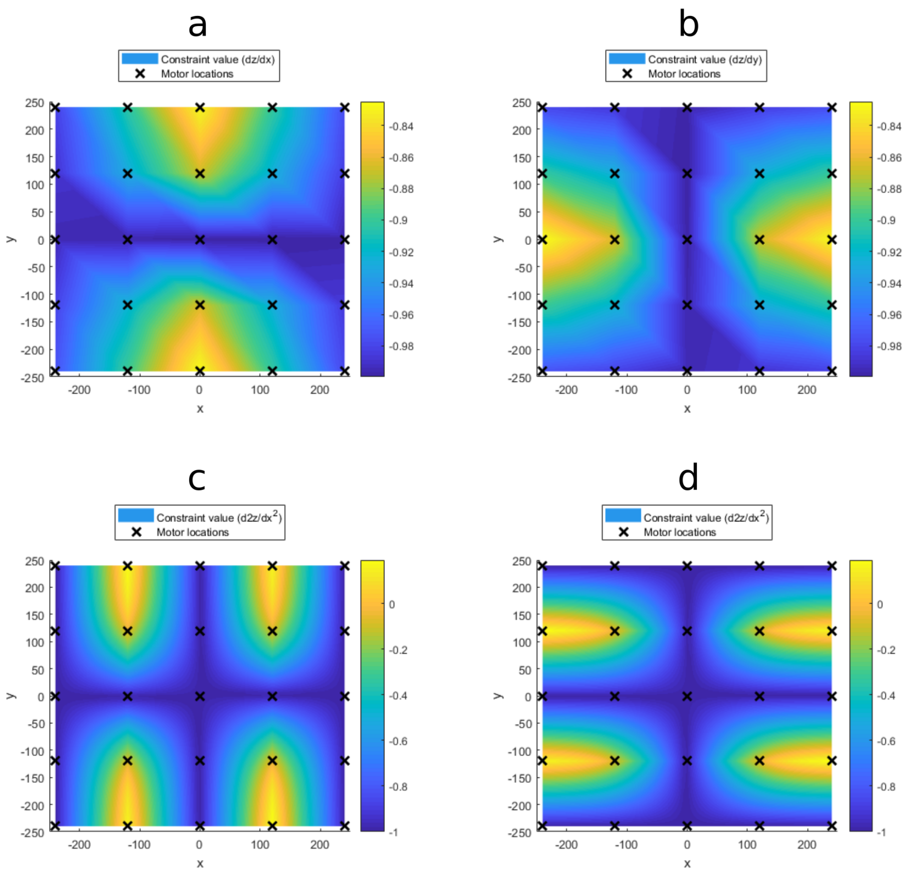

Fig. 4.5: The constraint values for the unconstrained solution. Figures a and b show the slope constraints, while c and d show the curvature constraints. Notice that values > 0 exist in c and d, indicating violated curvature constraints.

particular significance, but has been deliberately constructed such that the mould cannot reproduce the resulting shape exactly. This is done to illustrate the following important characteristic of working with adjustable moulds.

Adjustable moulds have a limited number of degrees of freedom (the number of actuators) for specifying the mould surface geometry. Therefore, only in very particular cases will it be possible to reproduce a given shape exactly, since the mould surface shape between actuators will be completely determined by the properties of the interpolating membrane. In general, reproducing a shape on such a system can therefore be done in different ways depending on requirements, through the selection of objective function. In the following the (weighted) mean-squared-error is used, which will yield a low average error.

The used mould has a square work surface with a side length of 480mm. Furthermore, it has physical limitations on the slope of the interpolating membrane above each actuator, and on the maximum curvature everywhere. For the present example, the linear approach described in section 4.2 is used, and the mould behaviour is modelled according to (4.1).

For reference, the unconstrained optimal mould shape is found. In Figures 4.2, 4.3, and 4.5 some of the plots that may be generated with the developed framework are shown. In Figure 4.2 the mould in the found optimal configuration (green surface) and defined panel (red surface) are shown. Figure 4.3 shows the mould surface shape error, while Figure 4.5 shows the values of the constraints (values > 0 indicate violated constraints). Notice that the curvature constraints are violated at multiple locations, see Figures 4.5c and 4.5d.

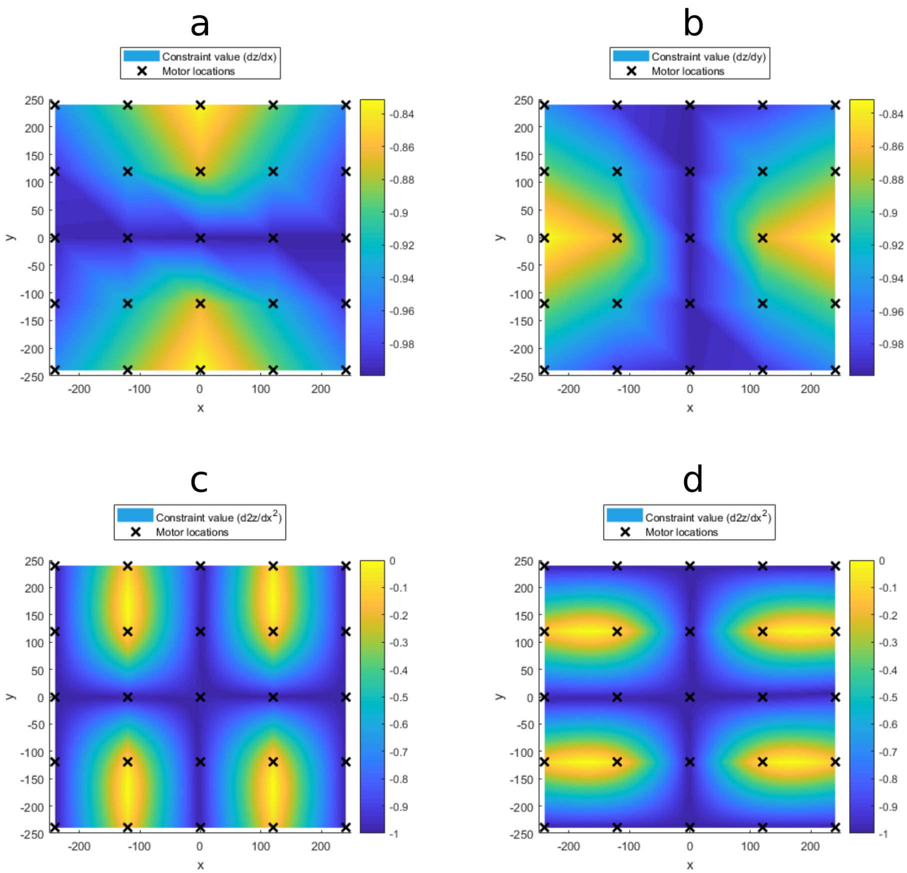

Fig. 4.6: The constraint values for the constrained solution. Figures a and b show the slope constraints, while c and d show the curvature constraints. Notice that the constraints are satisfied everywhere.

The mould constraints are now activated and a new solution is found. Figure 4.4 shows the error plot for this solution and Figure 4.6 shows the values of the constraints. Notice that all constraints are now satisfied, but that the mould surface shape error is now naturally worse than for the unconstrained case.

Finally, imagine that the shape accuracy of the upper right corner of the defined panel is of particular importance. In Figure 4.7 this corner is selected, and the weight of that area is increased to 100 (the default weight is 1). In Figure 4.8 an error plot for the weighted solution is shown, and in Figure 4.9 the corresponding constraint values are given. It is apparent that the shape error is now lower on average in the selected region, at the expense of overall higher error in the rest of the defined panel. As seen in Figure 4.9 the constraints are still satisfied everywhere.

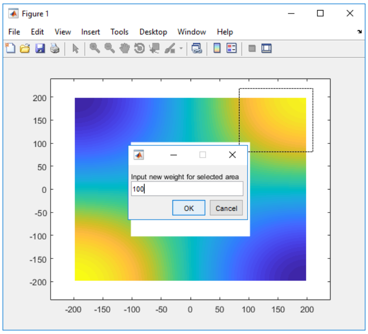

Fig. 4.7: A region of importance has been selected. The weights for this region are increased to 100.

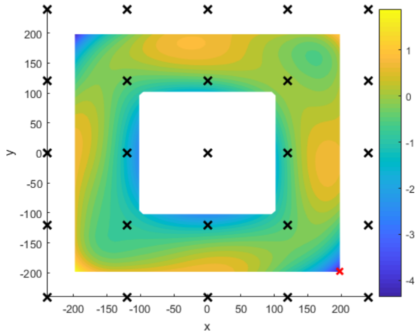

Fig. 4.8: An error plot showing the discrepancy between the predicted mould shape and the shape of the defined panel, for the constrained and weighted solution.

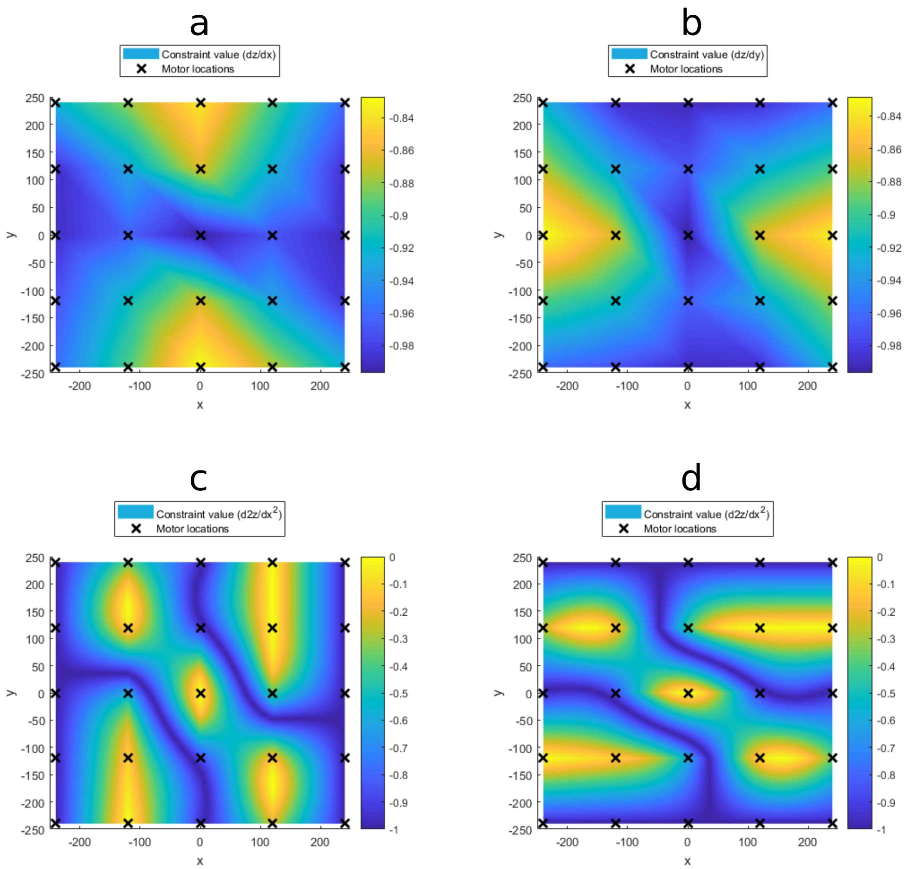

Fig. 4.9: The constraint values for the constrained and weighted solution. Figures a and b show the slope constraints, while c and d show the curvature constraints.

Chapter 5

Summary of Papers

In the following an introduction to the included papers is given. First the chronology and context of the work behind the three papers are briefly presented. Following this, a more detailed introduction to each paper is given, which emphasizes the motivations, objectives, and key findings.

5.1 Chronology and Context

For roughly the first year of the PhD project, the mechanical characterization of the AdaptiveMould was the focus research area. While not leading to the publication of any research papers, this work resulted in an in-depth understanding of the system that greatly benefited the continued work as well as Adapa's ongoing development of the system.

Researching the field of surrogate modelling was then commenced. About half way into the PhD a four month research stay at the University of Southampton, UK, under Dr. Alexander Forrester began. Prior to departure, the decision was made to base initial model development on the generic substitute model described in section 3.3.1. During the stay the ground work for Paper A was laid. A number of surrogate modelling techniques were examined and tested on the problem, and the focus was narrowed towards Kriging and proper orthogonal decomposition.

Upon returning from Southampton, model development continued while the experiments for obtaining measurement data were being planned. Simultaneously a paper was prepared for the 2015 ASME International Mechanical Engineering Congress and Exposition (IMECE). While this paper went through peer review during the following months, the work on obtaining the initial measurement data on a prototype AdaptiveMould was carried out. In November of 2015 IMECE was attended where the prepared paper and preliminary results based on the measurement data were presented. The IMECE paper unfortunately had to be pulled from publication due to concerns with respect to Adapa's intellectual property. The paper was therefore adapted, expanded, and published as a research paper. This paper was published in the ASME Journal of Computing and Information Science in Engineering, and is summarized in section 5.2 (paper A).

When benchmarking on the initial measurement data, it was clear that the developed surrogate models performed very poorly compared to the performance on data from the generic substitute model. When investigating the cause, it became clear that this discrepancy was due to a high amount of mechanical freeplay between the actuators and the interpolating membrane on the prototype used for the measurements. This introduced a non-linearity that was not accounted for in the generic substitute model. It was demonstrated that the character of this non-linearity made the prototype behave unpredictably. Based on this feedback Adapa developed a new actuator-to-membrane-interface with much improved tolerances, and thus minimizing mechanical freeplay.

A new set of measurements were planned for the improved AdaptiveMould prototype. Due to mechanical constraints in the AdaptiveMould systems (limitations on the slope of the interpolating membrane over the actuators and on the curvature of the membrane), sampling valid mould configurations with large actuator displacement ranges posed a problem. Therefore a novel sampling method - constrained orthogonal-maximin nested Latin hypercube sampling - was developed for this purpose. During the final year of the PhD, a second set of measurements were performed. A research paper concerning the findings with respect to mechanical freeplay, a benchmark of the developed surrogate models on measurement data before and after improvement of the actuator-to-membrane-interface, and a brief presentation of the developed sampling method was written during this period. This paper was published in the ASME Journal of Mechanical Design, and is summarized in section 5.3 (paper B).

Towards the end of the PhD project, further work was done to improve the constrained orthogonal-maximin nested Latin hypercube sampling method. A research paper detailing the improved algorithm and demonstrating its capabilities on simple examples as well as its application to the problem of the AdaptiveMould was written in the final months of the PhD. This paper is yet-to-be-published, and is summarized in section 5.4 (paper C).

5.2 Paper A: Developing Meta-Models for Fast and Accurate Prediction of the Draping of Physical Surfaces

The objective for paper A [33] was the development of one or more surrogate models for predicting the behaviour of the generic model presented in subsection 3.3.1 (although in a 9 by 9 actuator setup). In the paper, four different hybrid methods are benchmarked and compared on the problem. The treated methods utilize a difference structure [35] for combining a computationally efficient but inaccurate spline based model with Kriging [22] or with a combination of Kriging and proper orthogonal decomposition [38,39].

Using a numerical example, it is shown that the treated problem ideally requires a non-stationary method. Therefore, two of the treated models use a patch-wise modelling scheme, where a number of locally valid models work together, providing accurate predictions over the entire mould surface. This model structure allows discretely varying model parameters across the mould surface, essentially allowing a degree of non-stationarity.

Through a benchmark study it is shown, that the methods based on proper orthogonal decomposition perform very well on the treated problem, both with respect to accuracy and numerical efficiency. Furthermore it is shown that the employed difference structure is advantageous compared to simpler modelling approaches, and that the developed patch-wise model structure improves model accuracy.

5.3 Paper B: Experimental Validation of Surrogate Models for Predicting the Draping of Physical Interpolating Surfaces

The objective of paper B [34] was the experimental validation of the models developed in [33].

In [33] training and test data were obtained using straight forward Latin hypercube sampling. In order to do this, the range of movement of the mould actuators had to be severely limited to ensure that physical limitations of the system were not violated. In this paper this experimental limitation was solved through the development of a constrained orthogonal-maximin Latin hypercube sampling method. This sampling method was later further developed and was presented in paper C [40].

Using the developed sampling method and the straight forward sample plan used in [33], a total of 1350 mould surface geometries were measured using digital image correlation on two variable shape mould prototypes provided by Adapa. The models developed in [33] were benchmarked and compared using this data.

Through the comparison of the benchmark results, and by the help of an independent numerical study it was shown, that mechanical free-play in the actuator to membrane interface in a variable shape mould severely limits the accuracy of the model predictions. Using the numerical study, it was demonstrated that the reason is that this free-play results in additional non-linearities in the mould behaviour, resulting in a greatly increased need for training data.

The second variable shape mould prototype provided by Adapa was developed based on this knowledge, and hence featured greatly improved mechanical tolerances. Through the benchmark study carried out for the measurement data obtained from this prototype, the developed surrogate models were shown to perform well.

5.4 Paper C: Constrained Orthogonal-Maximin Nested Latin Hypercube Sampling

The objective of paper C [40] was the development of an algorithm for producing high quality sample designs, respecting some set of constraints. As mentioned, this sample method was needed for the work done in [34].

An algorithm for producing constrained orthogonal-maximin nested Latin hypercube sample designs is presented in the paper. Each term of this long title is clarified in the following.

A Latin hypercube sample design is stratified. This means that the sample design, if projected onto any axis of the space it exists in, the samples will be approximately uniformly distributed. In an experimental situation this ensures, that the entire range of possible values of a variable is exercised.

Such a design is not unique for a given dimensionality and number of samples. Therefore, it is common to select a good sample design through optimizing the sample design on a performance metric. The orthogonal-maximin [41] criterion is one such performance metric, that rewards low factor correlation and good space fillingness.

The term 'nested' refers to the fact that multiple sample designs exist as subsets of a single sample design. This property can be useful in multiple situations. For instance, in the case of co-kriging [31] or other multi-fidelity models, training data from multiple fidelities (resulting in different cost w.r.t. computational resources, human resources etc.) is combined to produce a high quality model. Such a model can be efficiently built using a nested sample design. The generation of random nested Latin hypercube sample designs is discussed in [42].

A constrained sample design refers to the fact that the individual sample locations in the design are subject to one or more constraints.

In the paper, the problem of finding a constrained orthogonal-maximin nested Latin hypercube sample design is formulated as a discrete Lagrange optimization problem, which is then solved using a simulated annealing stochastic optimizer [43]. The algorithm is demonstrated on two examples, and on the application of [34].

5.5 Contributions and Impact

In the literature on variable shape moulds, a lot of attention has been given to improving the mechanical designs and their applications. Efforts with respect to improving the quality of the products produced on such moulds have largely been concerned with improving the surface finish rather than shape accuracy.

In this work, shape accuracy has been the focus. The mechanical design of the mould system is indeed very important. However, it is important to note, that even in a system with perfect mechanical tolerances and an excellent mechanical design, the geometries that are possible to produce on the mould are determined by the behaviour of the interpolator. With accurate predictions of this behaviour, optimization can be used to achieve the best possible shape accuracy in use.

In paper A, an important contribution towards achieving such predictions is made. It is shown that it is possible to train a surrogate model that can provide accurate predictions of the behaviour of a system of this type in a computationally efficient way. The most promising of the methods presented in the paper is a hybrid method where a surrogate model based on Kriging and proper orthogonal decomposition corrects an inaccurate fast spline based model. In order to allow for non-stationarity of the model, a novel patch-wise modeling scheme is presented. This approach improves accuracy and provides better scalability since patch models can be reused across the mould surface.

Paper B expands on the work in paper A. In paper A, the developed models were trained and benchmarked on data from a numerical model with small actuator displacement ranges. Paper B contributes with an assessment of the performance of the developed models in a more realistic setup, where data comes from digital image correlation measurements on a physical mould, using much greater actuator displacement ranges. A further contribution of paper B is that it is explained why mechanical freeplay is so detrimental to model performances.

The mechanical constraints of the treated type of variable shape mould (limitations on membrane slope and curvature) made efficient sampling for training and test data generation difficult. Paper C presents a solution for this and similar problems. The developed algorithm for generating constrained orthogonal-maximin nested Latin hypercube sample designs provides an important and widely applicable contribution.

5.6 Suggestions for Future Work

The research carried out during the Ph.D. naturally leads to a number of questions or topics of interest for further research. In the following, some of these are discussed.

In papers A and B it was shown that the developed surrogate models were able to provide accurate predictions with good numerical efficiency. In paper B a performance assessment was performed for a small prototype variable shape mould with a square mould surface with a side length of 480mm. For a measurement volume of this size, digital image correlation already seemed impractical, due to the large distance needed between the mould surface and the cameras. For measurements on a full-scale mould research into other types of surface measurements could prove beneficial.

With respect to the patch-wise modelling scheme it has been argued that for large mould surfaces, model reuse across patches could possibly be employed. It was however also noted, that proximity to the borders of the mould surface has an influence on the behaviour. An examination of the extent to which this is the case, and of the distance from the border at which this effect is no longer significant would be needed to reliably employ model reuse. In this regard it would also be interesting to examine whether model reuse between similar on- or close-to-border patches is possible.

In this work the surrogate models have been developed disregarding potentially important effects, namely the effects of shape history and load. During the work it has been seen that the history of shapes that the mould has gone through has an effect on the current shape. It would be interesting to see if either this effect could be included in the developed models, or alternatively a 'reset-procedure' could be developed for the model to suppress this dependency. When the variable shape mould is in use, a specimen will be placed on the mould surface. This will result in loads on the mould surface from gravity, curing and thermal stress etc. This will naturally affect the behaviour of the mould surface shape. An examination regarding the effect of such a load should be carried out.

One of the research questions was concerned with the use of optimization in conjunction with the developed surrogate models, for determining the required actuator displacements for a given mould shape. A simple optimization approach using a linear prediction model instead of one of the developed models has been examined. An examination of the full approach using the most promising surrogate model should be carried out. Even though the developed models are numerically efficient to evaluate, the use of an optimization algorithm will require hundreds or even thousands of evaluations, resulting in significant run-time. Further research into improving the numerical efficiency of the complete system could prove beneficial. One possibility could be to examine if the gradient of the surrogate models under a given objective function could be formulated explicitly, so finite-difference gradient evaluation could be avoided during optimization.

References

[1] C. Munro and D. F. Walczyk, "Reconfigurable pin-type tooling: A survey of prior art and reduction to practice," *Journal of Manufacturing Science and Engineering*, vol. 129, no. 3, pp. 551–565, 2007, DOI: 10.1115/1.2714577.

[2] D. F. Walczyk and D. E. Hardt, "Design and analysis of reconfigurable discrete dies for sheet metal forming," *Journal of Manufacturing Systems*, vol. 17, no. 6, pp. 436–454, 1998, DOI: 10.1016/s0278-6125(99)80003-x.

[3] E. Haas, R. C. Schwarz, and J. M. Papazian, "Design and test of a reconfigurable forming die," *Journal of Manufacturing Processes*, vol. 4, no. 1, pp. 77–85, 2002, DOI: 10.1016/s1526-6125(02)70134-5.

[4] E. V. Sullivan, E. G. Haas, R. C. Schwarz, M. Kesselman, A. N. Peck, and J. M. Papazian, "Individual motor pin module," 2000, US Patent 6012314.

[5] D. E. Hardt, R. D. Webb, and N. P. Suh, "Sheet metal die forming using closed-loop shape control," *CIRP Annals - Manufacturing Technology*, vol. 31, no. 1, pp. 165–169, 1982, DOI: 10.1016/s0007-8506(07)63290-9.

[6] R. D. Webb and D. E. Hardt, "A transfer function description of sheet metal forming for process control," *Journal of Engineering for Industry*, vol. 113, no. 1, pp. 44–52, 1991, DOI: 10.1115/1.2899621.

[7] D. F. Walczyk, J. F. Hosford, and J. M. Papazian, "Using reconfigurable tooling and surface heating for incremental forming of composite aircraft parts," *Journal of Manufacturing Science and Engineering*, vol. 125, no. 2, pp. 333–343, apr 2003, DOI: 10.1115/1.1561456.

[8] Y. Wang, Z. J. Wang, and N. Gindy, "A method for representation of component geometry using discrete pin for reconfigurable moulds," *Advances in Engineering Software*, vol. 42, no. 7, pp. 409–418, 2011, DOI: 10.1016/j.advengsoft.2011.03.004.

[9] B. Koc and S. Thangaswamy, "Design and analysis of a reconfigurable discrete pin tooling system for molding of three-dimensional free-form objects," *Robotics and Computer-Integrated Manufacturing*, vol. 27, no. 2, pp. 335–348, apr 2011, DOI: 10.1016/j.rcim.2010.07.017.

[10] D. Rietbergen, "Adjustable mould for architectural freely curved glass," in *Proceedings of Challenging Glass*, 2008. [Online]. Available: http://www.glassfiles.com/articles/adjustable-mould-architectural-freely-curved-glass

[11] D. Rietbergen and K. J. Vollers, "A method and apparatus for forming a double-curved panel from a flat panel," 2009, PCT Patent Application WO 2009/002158 A1.

[12] J. Belis, A. D. C. Pronk, W. B. Schuurmans, and T. Blancke, "All-glass shell scale models made with an adjustable mould," in *Proceedings of the IABSE-IASS Symposium*, 2011. [Online]. Available: http://library.tue.nl/csp/vlink/VlinkMenu.CSP?sid=VubisSmart:TUE{&}{#}38;pid=oai:vubissmart:WEBOPACTUE:6:733076

[13] S. Grünewald, B. Janssen, H. R. Schipper, K. J. Vollers, and J. C. Walraven, "Deliberate deformation of concrete after casting," in *Proceedings of the Second International Conference on Flexible Formwork*, 2012, ISBN: 9780861971862. [Online]. Available: http://www.worldcat.org/isbn/9780861971862

[14] M. Kesselman and E. G. Haas, “Adjustable form die,” 1996, US Patent 5546784.

[15] M. K. Kristensen and K. R. Jepsen, “A flexible mat for providing a dynamically reconfigurable double-curved moulding surface in a mould,” 2012, PCT Patent Application WO 2012/065614 A1.

[16] S. A. Kalogirou, “Artificial neural networks in renewable energy systems applications: a review,” *Renewable and Sustainable Energy Reviews*, vol. 5, no. 4, pp. 373–401, dec 2001, DOI: 10.1016/S1364-0321(01)00006-5.

[17] J. C. Carr, W. R. Fright, and R. K. Beatson, “Surface interpolation with radial basis functions for medical imaging.” *IEEE transactions on medical imaging*, vol. 16, no. 1, pp. 96–107, feb 1997, DOI: 10.1109/42.552059.

[18] V. Klass, M. Behm, and G. Lindbergh, “Capturing lithium-ion battery dynamics with support vector machine-based battery model,” *Journal of Power Sources*, vol. 298, pp. 92–101, dec 2015, DOI: 10.1016/j.jpowsour.2015.08.036.

[19] X. Shi, Â. Palos Teixeira, J. Zhang, and C. Guedes Soares, “Kriging response surface reliability analysis of a ship-stiffened plate with initial imperfections,” *Structure and Infrastructure Engineering*, vol. 11, no. 11, pp. 1450–1465, nov 2014, DOI: 10.1080/15732479.2014.976575.

[20] M. Hamdaoui, F.-Z. Oujebbour, A. Habbal, P. Breitkopf, and P. Villon, “Kriging surrogates for evolutionary multi-objective optimization of CPU intensive sheet metal forming applications,” *International Journal of Material Forming*, vol. 8, no. 3, pp. 469–480, oct 2014, DOI: 10.1007/s12289-014-1190-y.

[21] D. C. Montgomery, *Design and Analysis of Experiments*, 8th ed. Wiley, 2012, ISBN: 9781118146927.

[22] A. I. J. Forrester and A. J. Keane, “Recent advances in surrogate-based optimization,” *Progress in Aerospace Sciences*, vol. 45, no. 1-3, pp. 50–79, 2009, DOI: 10.1016/j.paerosci.2008.11.001.

[23] A. Ng, “Cs229: Machine learning,” 2008, online course watched at https://www.youtube.com/course?list=ECA89DCFA6ADACE599, course homepage: http://cs229.stanford.edu/.

[24] J. H. Friedman, “Multivariate Adaptive Regression Splines,” *The Annals of Statistics*, vol. 19, no. 1, pp. 1–67, 1991, DOI: 10.1214/aos/1176347963.

[25] R. Jin, W. Chen, and T. W. Simpson, “Comparative studies of metamodelling techniques under multiple modelling criteria,” *Structural and Multidisciplinary Optimization*, vol. 23, no. 1, pp. 1–13, dec 2001, DOI: 10.1007/s00158-001-0160-4.

[26] D. J. C. MacKay, *Information theory, inference, and learning algorithms*, 4th ed. Cambridge University press, 2004, ISBN: 9780521642989.

[27] A. A. Mullur and A. Messac, “Extended Radial Basis Functions: More Flexible and Effective Metamodeling,” *AIAA Journal*, vol. 43, no. 6, pp. 1306–1315, 2005, DOI: 10.2514/1.11292.

[28] A. Smola and B. Schölkopf, “A tutorial on support vector regression,” *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004, DOI: 10.1023/B:STCO.0000035301.49549.88.

[29] S. M. Clarke, J. H. Griebsch, and T. W. Simpson, "Analysis of Support Vector Regression for Approximation of Complex Engineering Analyses," *Journal of Mechanical Design*, vol. 127, no. 6, 2005, DOI: 10.1115/1.1897403.

[30] C. E. Rasmussen, *Gaussian processes for machine learning*, 1st ed. MIT Press, 2006, ISBN: 026218253X.

[31] A. I. J. Forrester, A. Sóbester, and A. J. Keane, *Engineering Design via Surrogate Modelling: A Practical Guide*. Wiley, 2008, ISBN: 9780470060681.

[32] V. R. Joseph, Y. Hung, and A. Sudjianto, "Blind Kriging: A New Method for Developing Metamodels," *Journal of Mechanical Design*, vol. 130, no. 3, 2008, DOI: 10.1115/1.2829873.

[33] E. T. Christensen, A. I. J. Forrester, E. Lund, and E. Lindgaard, "Developing Meta-Models for Fast and Accurate Prediction of the Draping of Physical Surfaces," *Journal of Computing and Information Science in Engineering*, vol. 18, no. 2, 2018, DOI: 10.1115/1.4039334.

[34] E. T. Christensen, E. Lund, and E. Lindgaard, "Experimental Validation of Surrogate Models for Predicting the Draping of Physical Interpolating Surfaces," *Journal of Mechanical Design*, vol. 140, no. 1, 2018, DOI: 10.1115/1.4038073.

[35] J. E. Rayas-Sanchez, "EM-based optimization of microwave circuits using artificial neural networks: the state-of-the-art," *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, no. 1, pp. 420–435, 2004, DOI: 10.1109/tmtt.2003.820897.

[36] J. Zheng, X. Shao, L. Gao, P. Jiang, and H. Qiu, "A prior-knowledge input LSSVR metamodeling method with tuning based on cellular particle swarm optimization for engineering design," *Expert Systems with Applications*, vol. 41, no. 5, pp. 2111–2125, apr 2014, DOI: 10.1016/j.eswa.2013.09.010.

[37] T. Goel, R. T. Haftka, W. Shyy, and N. V. Queipo, "Ensemble of surrogates," *Structural and Multidisciplinary Optimization*, vol. 33, no. 3, pp. 199–216, 2007, DOI: 10.1007/s00158-006-0051-9.

[38] J. Shlens, "A Tutorial on Principal Component Analysis," *ArXiv e-prints*, apr 2014, arXiv:1404.1100 [cs.LG]. [Online]. Available: https://arxiv.org/abs/1404.1100

[39] M. Xiao, P. Breitkopf, R. Filomeno Coelho, C. Knopf-Lenoir, M. Sidorkiewicz, and P. Villon, "Model reduction by CPOD and Kriging," *Structural and Multidisciplinary Optimization*, vol. 41, no. 4, pp. 555–574, 2010, DOI: 10.1007/s00158-009-0434-9.

[40] E. T. Christensen, E. Lund, and E. Lindgaard, "Constrained Orthogonal-Maximin Nested Latin Hypercube Sampling," *To be submitted*.

[41] V. R. Joseph and H. Ying, "Orthogonal-Maximin Latin Hypercube Designs," *Statistica Sinica*, vol. 18, no. 1, pp. 171–186, 2008. [Online]. Available: https://www.jstor.org/stable/24308251

[42] P. Z. G. Qian, "Nested Latin hypercube designs," *Biometrika*, vol. 96, no. 4, pp. 957–970, dec 2009, DOI: 10.1093/biomet/asp045.

[43] B. Wah and T. Wang, "Constrained simulated annealing with applications in nonlinear continuous constrained global optimization," in *Proceedings 11th International Conference on Tools with Artificial Intelligence*. IEEE Comput. Soc, 1999, pp. 381–388, DOI: 10.1109/TAI.1999.809827.

Part II

Papers

Paper A

Developing Meta-Models for Fast and Accurate Prediction of the Draping of Physical Surfaces

Esben T. Christensen[1], Alexander I. J. Forrester[2], Erik Lund[1], and Esben Lindgaard[1]

[1]*Department of Materials and Production, Aalborg University, Fibigerstræde 16, DK-9220 Aalborg Øst, Denmark*
[2]*Engineering Centre of Excellence, University of Southampton, Burgess Road, SO16 7QF Southampton, United Kingdom*

The paper has been published in the
Journal of Computing and Information Science in Engineering
Vol. 18(2), 2018.

Paper B

Experimental Validation of Surrogate Models for Predicting the Draping of Physical Interpolating Surfaces

Esben T. Christensen[1], Erik Lund[1], and Esben Lindgaard[1]

[1]*Department of Materials and Production, Aalborg University, Fibigerstræde 16, DK-9220 Aalborg Øst, Denmark*

The paper has been published in the
Journal of Mechanical Design
Vol. 140(1), 2018.

Paper C

Constrained Orthogonal-Maximin Nested Latin Hypercube Sampling

Esben T. Christensen[1], Erik Lund[1], and Esben Lindgaard[1]

[1]*Department of Materials and Production, Aalborg University, Fibigerstræde 16, DK-9220 Aalborg Øst, Denmark*

To be published...

Constrained Orthogonal-Maximin Nested Latin Hypercube Sampling

Esben Toke Christensen[a]*, Erik Lund[a] and Esben Lindgaard[a]

[a]*Department of Materials and Production, Aalborg University, Aalborg, Denmark*

*esben@mp.aau.dk

Constrained Orthogonal-Maximin Nested Latin Hypercube Sampling

Latin hypercube sampling is a popular methodology across a wide range of disciplines. Often design constraints is an important aspect of a sampling strategy but this has only received little attention in connection with Latin hypercube sampling. In many applications, the further requirement of nested sample designs is of particular importance. In this paper, an algorithm for producing constrained orthogonal-maximin nested Latin hypercube sample designs is presented. The algorithm comprises the creation of a random Latin hypercube sample design, followed by the numerical solution of a discrete Lagrangian optimization problem using simulated annealing. The algorithm and the designs it produces are demonstrated on two examples and an application.

Keywords: nested Latin hypercube sampling; discrete Lagrange problem; optimization; design of experiments; constrained sampling; simulated annealing

Introduction

Since it was introduced (McKay *et al.* 1979) Latin hypercube (LH) sampling has been a popular choice of sampling methodology in a wide range of applications, particularly in surrogate- or metamodeling, see e.g. Forrester and Keane (2009).

In surrogate modeling a data-driven model of a system (or computationally expensive simulation) is established based on a sample design that specifies a number of configurations of the independent variables of the system. In the particular case of surrogate-based optimization, a surrogate model that predicts the value of the (expensive) objective function from the independent variables is often initialized in this way. Optimization is then carried out on this model, only occasionally updating the surrogate model through evaluation of the underlying true objective function. Surrogate-based optimization can in many problems massively decrease the time needed for finding an optimum.

The accuracy of a surrogate model, and in turn the effectiveness of surrogate-based optimization, can be very dependent on the sampling scheme used. Latin hypercube sampling has been a successful choice in numerous applications due to its advantageous properties.

One of the most notable advantages of a general Latin hypercube sample design is that the samples are stratified and uniformly distributed if projected onto any axis of the hypercube. For a given number of samples and dimensionality many such Latin hypercube designs exist, most of which will have poor properties. Typically, to obtain LH sample designs with desirable properties, a design which optimizes a performance metric that reflects these properties is selected. In many cases, a sample design consisting of several sample designs nested inside each other is advantageous. Notable examples include the construction of multi-fidelity models such as co-kriging (Forrester and Keane 2009) and benchmarking scenarios (Christensen, Forrester, *et al.* 2018, Christensen, Lund, *et al.* 2018). Often a sample design must furthermore satisfy a set of constraints. A method for finding a sample design which meets the mentioned requirements, namely a nested Latin hypercube sample design which satisfies a set of constraints and has good properties, is warranted.

Much work has previously been conducted with respect to quantifying the quality of a given sample design. The maximin criterion (Morris and Mitchell 1995) rewards designs where the minimum distance between any two samples is large. Finding a design which scores well under this criterion therefore amounts to finding a space filling design. Owen (1994) instead suggests ranking sample designs with a criterion that rewards low correlation between the factors of the design. In Joseph and Ying (2008) a criterion termed the orthogonal-maximin criterion is proposed. This is a weighted criterion which compounds the former two. Selecting a sample design based on this criterion therefore amounts to finding a design where both factor correlation and space fillingness is considered.

The actual task of finding a good Latin hypercube sample design according to a particular performance metric is solved as an optimization problem. This problem is difficult to solve since the space of Latin hypercube designs is typically huge, due to being combinatorial in nature, and the behavior of the performance metric is complicated with a large number of local minima. Typically, the problem is tackled by seeking to improve the criterion by iteratively manipulating a randomly initialized design using e.g. a simulated annealing algorithm. In Joseph and Ying (2008) such an approach is presented. The further complication that follows from the introduction of constraints on a Latin hypercube sample design has not been given much attention. Petelet *et al.* (2010) provides an approach for handling the specific case where inequality constraints between the sampled factors exist, i.e. $x_1 > x_2$ and so on. For more general scenarios, e.g. more general sets of linear constraints or non-linear constraints, no literature has been found.

Several ways of generating nested sample designs exist. Perhaps the simplest way is generating a conventional sample design, and then selecting a nested design as a subset of this (Forrester *et al.* 2008). In a more direct approach Qian (2009) has introduced a method for generating random nested Latin hypercube sample designs, which is a more efficient approach that ensures that each nested design becomes a Latin hypercube design.

In this work a novel method for constructing high quality nested Latin hypercubes satisfying general linear or non-linear equality- and/or inequality constraints is suggested. The algorithm starts with a random nested Latin hypercube generated according to Qian (2009). In order to improve the quality of this design and resolve violated constraints, a discrete Lagrangian optimization problem is formulated and solved using a simulated annealing process (Wah and Wang 1999) followed by a gradient based optimizer. The quality of the design is quantified using the orthogonal-maximin criterion presented in Joseph and Ying (2008).

The paper is organized as follows: First the optimization problem and its discrete Lagrangian form is given, and the general structure of the algorithm is presented. Following this, the method for constructing a random nested Latin hypercube and the formulation of the sample design quality metric are summarized. The remaining specifics of the algorithm are then presented. Finally, two examples and an application demonstrating the algorithm are given, followed by a conclusion.

Constrained Orthogonal-Maximin Nested Latin Hypercubes

Consider the problem of finding a nested sample design, $\boldsymbol{X}$, of N samples in a k dimensional constrained sample space. This task can be stated in terms of the following optimization problem

$$\underset{\boldsymbol{X}}{\text{Minimize}}\big(\phi(\boldsymbol{X})\big)$$
$$\text{s.t. } \boldsymbol{h}(\boldsymbol{X}) = \begin{bmatrix} \tilde{\boldsymbol{h}}(\boldsymbol{X}_1) \\ \tilde{\boldsymbol{h}}(\boldsymbol{X}_2) \\ \vdots \\ \tilde{\boldsymbol{h}}(\boldsymbol{X}_N) \end{bmatrix} = \boldsymbol{0} \tag{1}$$

where ϕ is a performance metric measuring the quality of a nested Latin Hypercube (NLH) sample design, and $\tilde{\boldsymbol{h}}$ is a vector function of equality constraints. $\boldsymbol{X}_i$ should be understood as the ith sample/row of $\boldsymbol{X}$ and $\tilde{\boldsymbol{h}}$ are therefore constraints formulated for the sample space and should be satisfied by every point of a sample design. Without loss of generality, inequality constraints can be disregarded since any inequality constraint, $g(\boldsymbol{X}) \leq 0$, can be reformulated to an equality constraint as $\max(g(\boldsymbol{X}), 0) = 0$.

To solve the optimization problem, Eq. (1) is restated in the form

$$\underset{\boldsymbol{\lambda}}{\text{maximize}} \left(\inf_{\boldsymbol{X}} \big(L(\boldsymbol{X}, \boldsymbol{\lambda}) \big) \right) \tag{2}$$

with the generalized discrete Lagrangian function

$$L(\boldsymbol{X}, \boldsymbol{\lambda}) = \phi(\boldsymbol{X}) + \boldsymbol{\lambda}^{\mathrm{T}} |\boldsymbol{h}(\boldsymbol{X})| \tag{3}$$

as suggested in Wah and Wang (1999). The reason the problem is treated as a discrete problem will be elaborated later. Finding candidate constrained minima then amounts to finding saddle points of the unconstrained Lagrange function in Eq. (3), where L has a maximum w.r.t. $\boldsymbol{\lambda}$ and a minimum w.r.t $\boldsymbol{X}$. An algorithm for producing constrained orthogonal-maximin nested Latin hypercubes through the solution of Eq. (2) is proposed in Table 1. The algorithm is inspired by Morris and Mitchell (1995) and Wah and Wang (1999) and is very similar to common simulated annealing approaches. In fact, if lines 7 to 11 (Table 1) are removed, the algorithm principally equals the simulated annealing procedure used for improving the quality of Latin hypercube designs in Morris and Mitchell (1995), Joseph and Ying (2008). However, instead of purely minimizing the objective, in each iteration Eq. (3) is either sought maximized by altering the Lagrange multipliers with probability β (lines 7 through 10), or sought minimized by altering $\boldsymbol{X}$ with probability $1 - \beta$ (lines 11 to 14). In lines 15 and 16 the current design is stored as the best design observed so far, if improvement is observed. Following the simulated annealing, a gradient based optimizer is used for additional improvement in line 22. In the following the individual components and parameters of the algorithm are explained.

ALGORITHM

1. User input data:
   a. $n_0, \boldsymbol{t}, k, \boldsymbol{b}_{lower}, \boldsymbol{b}_{upper}$ — Specification of the problem.
   b. $p, w, \alpha, \beta_0, \eta, \epsilon, \epsilon_{Accept}, N_\epsilon$ — Algorithm parameters.

2. Define:
   a. Function $\boldsymbol{X} = \mathrm{NLH}(n_0, \boldsymbol{t}, k, \boldsymbol{b}_{lower}, \boldsymbol{b}_{upper})$ generates a random nested Latin hypercube sample design according to Qian (2009).
   b. Procedure $\boldsymbol{X}^* = \mathrm{G_X}(\boldsymbol{X})$ generates a new sample design $\boldsymbol{X}^*$ in the neighborhood of $\boldsymbol{X}$ preserving the NLH structure.
   c. Procedure $\boldsymbol{\lambda}^*, \boldsymbol{d\lambda}^* = \mathrm{G}_\lambda(\boldsymbol{\lambda}, \boldsymbol{d\lambda})$ generates a new vector of Lagrange multipliers $\boldsymbol{\lambda}^*$ in the neighborhood of $\boldsymbol{\lambda}$ and updates the allowed rates of change $\boldsymbol{d\lambda}^*$.
   d. Function $\phi = \phi(\boldsymbol{X})$ calculates the orthogonal-maximin criterion of $\boldsymbol{X}$ according to Joseph and Ying (2008).
   e. Function $\boldsymbol{X}^* = \mathrm{GBO}(\boldsymbol{X})$ determines the $\boldsymbol{d\pi}$ that maximizes $\phi(\boldsymbol{X})$ while respecting constraints using a gradient based optimizer.

3. Initialization:
   a. Select an initial NLH sample design as $\boldsymbol{X}_0 = \mathrm{NLH}(N_0, \boldsymbol{t}, k)$
   b. Initialize Lagrange multipliers $\boldsymbol{\lambda} = \boldsymbol{1}$ and neighborhood size $\boldsymbol{d\lambda} = \boldsymbol{1}/\max(|\boldsymbol{h}_0|)$.
   c. Set $\boldsymbol{X}_{best} = \boldsymbol{X}_0$ and $\boldsymbol{X} = \boldsymbol{X}_0$.
   d. Set T to appropriate start temperature, $T = T_0$.

```
4.  while stopping condition not satisfied do
5.      while equilibrium condition not satisfied do
6.          for I_T ← 1 to n_T do
7.              with probability β do
8.                  set λ_try, dλ = G_λ(λ, dλ)
9.                  if L(X, λ_try) > L(X, λ) or with probability e^(−(L(X,λ)−L(X,λ_try))/T)
10.                     set λ = λ_try
11.             otherwise
12.                 set X_try = G_X(X)
13.                 if L(X_try, λ) < L(X, λ) or with probability e^(−(L(X_try,λ)−L(X,λ))/T)
14.                     set X = X_try
15.                 if L(X_try, λ) < L(X_best, λ)
16.                     set X_best = X_try
17.             end
18.         end for
19.     end while
20.     set T = αT
21. end while
22. set X_best = GBO(X_best)
23. Return X_best
```

Table 1. Algorithm producing constrained orthogonal-maximin nested Latin hypercubes.

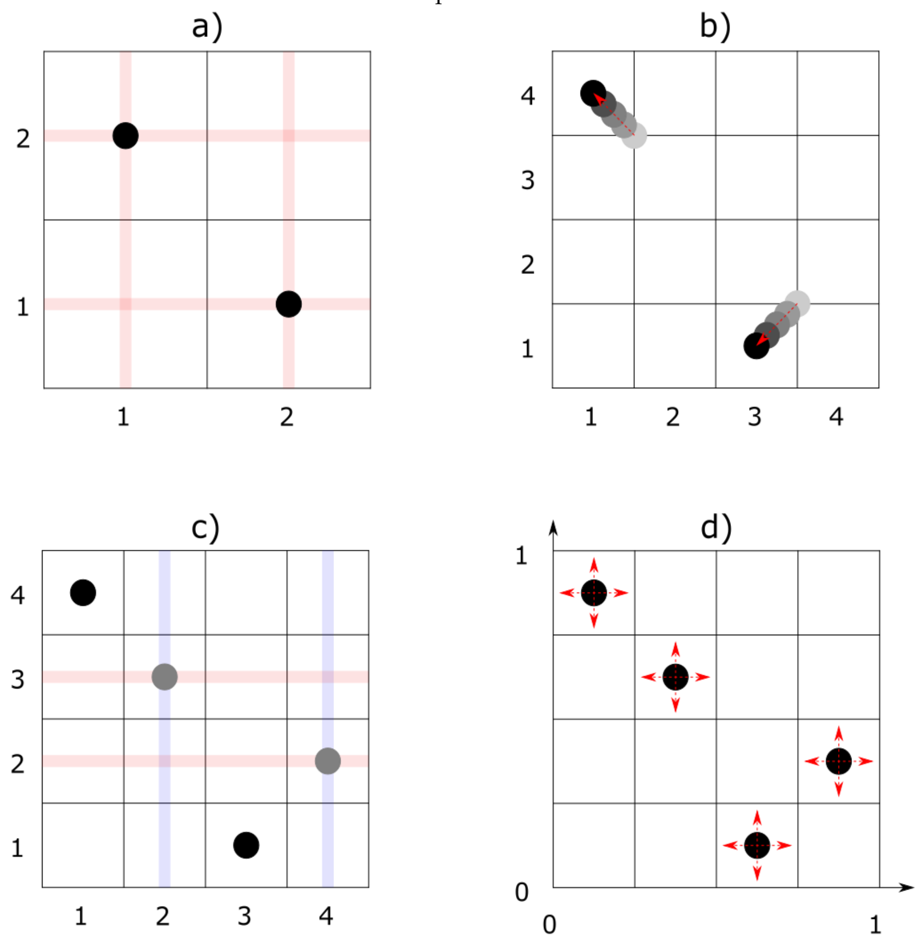

Figure 1. Illustration of the methodology used when constructing a nested Latin hypercube sample design in two layers of a total of four samples. a) An initial Latin hypercube of two samples in a two by two discretization is established. The red lines illustrate that any given sample is the only sample in its column and row. b) The design is expanded to a four by four discretization. To do this, each sample established in a) is moved randomly to one of the four cells surrounding it. c) The red and blue lines mark the now empty rows and columns, respectively. Their intersections correspond to valid locations for placing additional samples. To complete the nested sample design on the four by four discretization, two additional samples are placed such that every row and every column contains exactly one sample. On this discretization, the full black samples correspond to the first level sample design, while all four samples constitute the second level sample design. d) Finally, the discrete sample design can optionally be mapped to a continuous sample space, and each sample can be adjusted within its cell.

Generating a random initial Latin hypercube sample design

In order to produce a nested Latin hypercube sample design, the method presented in Qian (2009) is used. The method is easily understood through an example. Denote the number of samples in layer i by n_i and assume that the number of samples in layer i is an integer multiple, t_i, of the number of samples in layer $i-1$. Furthermore, let n_0 be given and let k be the number of factors in the design. For the purpose of the following example, set $k=2$, $n_0=2$ and $\boldsymbol{t}=[1,2]$ such that the number of samples in a two-layer design becomes $n_1=n_0t_1=2$ and $N=n_2=n_1t_2=4$.

Denote by Z_q the set of integers from 1 to q where q is some integer, $q \geq 1$. An initial interim discrete LH design, $\mathbf{\Pi}$, is generated by setting each column to a random permutation of Z_{n_1}. In the present example suppose this results in

$$\mathbf{\Pi} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \qquad (4)$$

This is illustrated in Figure 1a. For each additional layer in the nested structure the design is expanded through the following steps.

1) Define an interim LH design $\mathbf{T}$ for $i=2$ as

$$\mathbf{T}_{j,k} = U_{dis}\left[\left(\mathbf{\Pi}_{j,k}-1\right)\boldsymbol{t}_i+1, \mathbf{\Pi}_{j,k}\boldsymbol{t}_i\right] \qquad (5)$$

where $U_{dis}[a,b]$ produces a discrete uniform sample on the interval $[a,b]$. In the present example, suppose this results in

$$\mathbf{T} = \begin{bmatrix} 1 & 4 \\ 3 & 1 \end{bmatrix} \qquad (6)$$

where for example $\mathbf{T}_{1,2} = U_{dis}[(2-1)\cdot 2+1, 2\cdot 2] = U_{dis}[3,4] = 4$. This design can be seen as an approximation of $\mathbf{\Pi}$ as represented on a finer grid, where each factor is now partitioned into n_i levels. This step is illustrated in Figure 1b.

2) Let the operation $A\backslash B$ perform a removal of the intersection of the sets A and B from A. Define $\mathbf{P}$ where each column, $\mathbf{P}_j$, is a random permutation on $Z_{n_i}\backslash\mathbf{T}_j$. In the considered example, this could result in

$$\mathbf{P} = \left[\begin{bmatrix} 2\\4\\1\\3 \end{bmatrix} \backslash \begin{bmatrix} 1\\3 \end{bmatrix} ; \begin{bmatrix} 3\\1\\2\\4 \end{bmatrix} \backslash \begin{bmatrix} 4\\1 \end{bmatrix}\right] = \begin{bmatrix} 2 & 3\\4 & 2 \end{bmatrix} \tag{7}$$

Here $\mathbf{P}$ should be seen as one possible set of points that, together with $\mathbf{T}$, defines a complete Latin hypercube design on the present grid, where each factor is partitioned into n_i levels.

3) $\mathbf{\Pi}$ is updated by stacking $\mathbf{T}$ and $\mathbf{P}$. For the treated example, the resulting discrete NLH design then becomes

$$\mathbf{\Pi} = \begin{bmatrix} \mathbf{T}\\ \mathbf{P} \end{bmatrix} = \begin{bmatrix} 1 & 4\\3 & 1\\2 & 3\\4 & 2 \end{bmatrix} \tag{8}$$

Steps 2 and 3 of this procedure are illustrated in Figure 1c.

To expand to additional levels of the sample design, the described steps are simply repeated. After performing the described steps, $\mathbf{\Pi}$ constitutes the discrete nested Latin hypercube design. The individual designs that it contains are recovered as partitions of $\mathbf{\Pi}$, where $\mathbf{\Pi}^{(1)}$ consists of the first n_1 samples, $\mathbf{\Pi}^{(2)}$ of the first n_2 samples and so on. For the present example, the contained designs are

$$\mathbf{\Pi}^{(1)} = \begin{bmatrix} 1 & 4 \\ 3 & 1 \end{bmatrix} \text{ and } \mathbf{\Pi}^{(2)} = \begin{bmatrix} 1 & 4 \\ 3 & 1 \\ 2 & 3 \\ 4 & 2 \end{bmatrix} \tag{9}$$

The discrete design resulting from the above procedure can be converted to a continuous NLH design on the k dimensional unit hypercube as

$$\widetilde{\boldsymbol{X}}_{j,k} = (\mathbf{\Pi}_{j,k} - \boldsymbol{d}\mathbf{\Pi}_{j,k})/N \tag{10}$$

where $\boldsymbol{d}\mathbf{\Pi}_{j,k} \in [0,1]$ for all j, k and N is the total number of samples in $\mathbf{\Pi}$. During the simulated annealing $\boldsymbol{d}\mathbf{\Pi}_{j,k} = 0.5$ for all j and k, so that each sample is centered in its respective cell. During the gradient based optimization (line 22 in Table 1) $\boldsymbol{d}\mathbf{\Pi}$ is varied to improve the final quality of the sample design. This final step is illustrated in Figure 1d.

Following the optimization procedure each factor (column), $\widetilde{\boldsymbol{X}}_j$, of the design can be scaled to the particular problem at hand, simply as

$$\boldsymbol{X}_j = \widetilde{\boldsymbol{X}}_j\left(\boldsymbol{b}_{upper,j} - \boldsymbol{b}_{lower,j}\right) + \boldsymbol{b}_{lower,j} \tag{11}$$

where $\boldsymbol{b}_{upper}$ and $\boldsymbol{b}_{lower}$ are vectors of upper and lower bounds.

The performance metric

The metric, ϕ, used for evaluating the quality of a particular NLH sample design is the criterion suggested in Joseph and Ying (2008). This criterion has the form

$$\phi(\boldsymbol{X}) = w\phi_c(\boldsymbol{X}) + (1-w)\frac{\phi_d(\boldsymbol{X}) - \phi_{d,L}}{\phi_{d,U} - \phi_{d,L}} \tag{12}$$

and is a weighted criterion that rewards both low factor correlation and a large minimum inter-point distance. The parameter w is a user input weight in [0,1] that determines the relative importance of the factor correlation term, ϕ_c. Unless specific needs dictate otherwise $w = 0.5$ is suggested. The factor correlation term, ϕ_c, is given in Joseph and Ying (2008) as

$$\phi_c(\boldsymbol{X}) = \frac{\sum_{i=2}^{k}\sum_{j=1}^{i-1}\rho_{ij}^2}{k(k-1)/2} \tag{13}$$

where ρ_{ij} is the Pearson correlation between factors i and j, and k is the number of factors in the design. The value of ϕ_c can take values in the interval [0,1]. ϕ_d is the Morris-Mitchell (Morris and Mitchell 1995) criterion

$$\phi_d(\boldsymbol{X}) = \left(\sum_{i=1}^{\binom{N}{2}} \frac{1}{d_i^p}\right)^{\frac{1}{p}} \tag{14}$$

where the d_i are the ‘N choose 2’ inter-point distances. $\phi_{d,L}$ and $\phi_{d,U}$ are the lower and upper bounds of ϕ_d, respectively. For a discrete design with factor levels in Z_N, as exemplified in Eq. (8), Joseph and Ying (2008) gives the upper and lower bounds

$$\tilde{\phi}_{d,L} = \left(\binom{N}{2}\left(\frac{\lceil \bar{d}\,\rceil - \bar{d}}{\lfloor \bar{d}\,\rfloor^p} + \frac{\bar{d} - \lfloor \bar{d}\,\rfloor}{\lceil \bar{d}\,\rceil^p}\right)\right)^{\frac{1}{p}} \tag{15}$$

$$\tilde{\phi}_{d,U} = \left(\sum_{i=1}^{N-1} \frac{N-i}{(ik)^p}\right)^{\frac{1}{p}} \tag{16}$$

where $\bar{d}$ is the mean inter-point distance, and the operators $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ round down or up to nearest integer, respectively. As suggested in Joseph and Ying (2008) the L_1 inter-point distance metric can be used, in which case

$$\bar{d} = \frac{(N+1)k}{3} \qquad (17)$$

Since a continuous design with factor levels in $[0,1]$ is of interest, the estimates

$$\phi_{d,L} = \tilde{\phi}_{d,L}/N \qquad (18)$$

$$\phi_{d,U} = \tilde{\phi}_{d,U}/N \qquad (19)$$

are used in Eq. (12). The parameter p is input by the user and with a sufficiently high value, the design that minimizes ϕ_d is a maximin design. In the examples in this paper $p = 5$ is used. Using the lower and upper bounds, the Morris-Mitchell criterion is normalized to $[0,1]$ in Eq. (12), which ensures predictable behavior w.r.t. the weight w. With $w = 0.5$ the criterion in Eq. (12) therefore reflects factor correlation and minimum inter-point distance with equal emphasis.

Perturbing the sample design

A perturbation which preserves the structure of a Latin hypercube sample design can be performed by interchanging any two elements of a single column of the design, $\boldsymbol{X}$ (Joseph and Ying 2008). To also preserve the *nested* structure, the two elements to be interchanged must be selected within the same layer in the design. Generating a trial NLH sample design, $\boldsymbol{X}_{try} = G_X(\boldsymbol{X})$, is therefore done in the following way. First a factor (column) index, c^*, is chosen at random from a discrete uniform distribution. Next a sample (row) index, r^*, of $\boldsymbol{X}$ is chosen at random from a discrete probability distribution with probabilities $\boldsymbol{\eta}_1$. The layer number that the selected row belongs to, l, is found as the l for which $n_{l-1} < r^* \leq n_l$. Another row index, r', is then picked at random with probabilities $\boldsymbol{\eta}_2$ from the set $Z_l \backslash \{Z_{l-1} \cup r^*\}$ (the rows belonging to the same layer as r^* which are not part of any lower level layer). The trial design is then established by interchanging elements (r^*, c^*) and (r', c^*) of $\boldsymbol{X}$. Since $G_X(\boldsymbol{X})$ is defined in terms of interchanging column elements, the optimization problem becomes discrete.

The probabilities $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ are set so that the probability of selecting a particular row is proportional to the sum of the absolute constraint violations of that row. However, a small but non-zero probability is assigned to rows with no constraint violations, to prevent the transition probability matrix of the annealing process from becoming reducible. A non-reducible transition probability matrix ensures that every state in the entire state space of the problem can be reached with a non-zero probability, which is a requirement for ensuring global convergence of the simulated annealing algorithm (Lundy and Mees 1986). By assigning a higher probability to rows with high constraint violation, the algorithm will effectively spend more time trying to resolve those constraint violations.

Perturbing the Lagrange multipliers

Perturbation of the Lagrange multipliers is performed as suggested in Wah and Wang (1999) as

$$\boldsymbol{\lambda}_{try} = G_\lambda(\boldsymbol{\lambda}) = \boldsymbol{\lambda} + u[-1,1]\boldsymbol{d\lambda}_j \boldsymbol{e}^{(j)} \qquad (20)$$

where $\boldsymbol{d\lambda}_j$ is the j'th entry of a vector, $\boldsymbol{d\lambda}$**,** of maximal allowed changes to $\boldsymbol{\lambda}$, $\boldsymbol{e}^{(j)}$ is a vector of the same length as $\boldsymbol{\lambda}$, whose only non-zero element is a unit value in position j, and $u[-1,1]$ produces a uniformly distributed continuous random sample between -1 and 1. The index j of the Lagrange multiplier to be changed is selected at random from a uniform distribution over the currently active constraints.

Before every perturbation of a Lagrange multiplier, the allowed rate of change of that multiplier is updated as suggested in Wah and Wang (1999) as

$$d\lambda_j = \begin{cases} d\lambda_j \cdot \eta & \text{if} \quad |h_j(\boldsymbol{X})| > \tau \cdot T \\ d\lambda_j \cdot \dfrac{1}{\eta} & \text{if} \quad 0 < |h_j(\boldsymbol{X})| < \tau \cdot T \end{cases} \tag{21}$$

where $\eta > 1$, and $\tau > 0$ are constants, and T is the current temperature. In the examples presented in this paper $\eta = 1.5$ has been used, while $\tau = \max(|\boldsymbol{h}^{(0)}|)/T_0$ where T_0 is the initial temperature and $\boldsymbol{h}^{(0)}$ are the values of the constraints at $T = T_0$. The motivation behind Eq. (21) is, that if constraint $\boldsymbol{h}_j$ is decreasing too slowly the corresponding $\boldsymbol{d\lambda}_j$, which determines the allowed rate of change of the Lagrange multiplier, should be increased and vice versa. Initialization of $\boldsymbol{d\lambda}$ before starting the annealing is performed as

$$\boldsymbol{d\lambda}_i = \frac{1}{\max(|\boldsymbol{h}^{(0)}|)} \quad \text{for} \quad i = 1 \dots n_c \tag{22}$$

Additional details

In Wah and Wang (1999) a fixed value of $\beta = 0.05$ (see Table 1) is suggested. However, experience has shown that this has resulted in too frequent updates of the Lagrange multipliers when only a few constraints remain violated. Instead the following heuristic has proven successful. Denote the number of violated constraints by n_h and the equivalent number of variables of the NLH by n_e. For each iteration, the value of β is updated to

$$\beta = \min\left(\frac{n_h}{n_e}, \beta_0\right) \tag{23}$$

where $\beta_0 = 0.05$. The equivalent number of variables should be understood as the number of variables needed for specifying a NLH. An equivalent number is used since perturbations of the NLH are performed procedurally. A suggested estimate is $n_e = Nk$.

In the developed implementation the algorithm is stopped when the ratio of accepted perturbations of the design drops below a certain tolerance, ϵ_{Accept}, and the maximum change in the best objective value, $L(\boldsymbol{X}_{best}, \boldsymbol{\lambda})$, over the last N_ϵ temperatures is below a certain tolerance, ϵ ($\epsilon_{Accept} = 10^{-2}$, $N_\epsilon = 10$, and $\epsilon = 10^{-2}$ are used in the examples). If the algorithm finishes with unresolved constraints, reannealing with a slower cooling schedule is suggested. One should of course consider, if a feasible solution exists or if the constraints are too restrictive.

The loop starting on line 5 in Table 1 continues until an equilibrium at the current temperature is achieved. The process is considered in equilibrium when the mean of $L(\boldsymbol{X}_i, \mathbf{1})$ over the states $\boldsymbol{X}_i$ visited at the current temperature has stabilized. In practice, this is done by recording this mean every n_T iterations and keeping a history of the last N_ϵ means. When the relative difference between the largest and smallest mean value in the history drops below ϵ, the process is considered to have reached equilibrium, and the temperature is lowered as $T = \alpha T$. In the examples $n_T = Nk$ and $\alpha = 0.95$ has been used.

The method of setting the initial temperature, T_0, is inspired by Wah and Wang (1999). A number of design perturbations, $\boldsymbol{X}^{*(i)} = G_X(\boldsymbol{X}^{(0)})$, are generated from the initial design $\boldsymbol{X}^{(0)}$. The initial temperature is then selected as

$$T_0 = \max_{\mathrm{i,j}}\left(\left|L\left(\boldsymbol{X}^{*(j)}, \mathbf{1}\right) - L\left(\boldsymbol{X}^{(\mathbf{0})}, \mathbf{1}\right)\right|, \left|h_i\left(\boldsymbol{X}^{*(j)}\right)\right|\right) \tag{24}$$

as this will result in a large fraction of accepted trial designs at the beginning of the annealing process. In the examples treated in this paper T_0 is set using 10000 perturbations from $\boldsymbol{X}^{(0)}$.

Examples

Three examples illustrating the use of the suggested algorithm are given in the following. The first two examples serve mainly to visually illustrate that the algorithm works as intended, while the third example seeks to illustrate performance on a real problem.

A small two-dimensional problem

A two-dimensional nested Latin hypercube with two layers of 10 and 20 samples, respectively, is produced on the unit square. Each sample (row), $\boldsymbol{x}$, of the design must satisfy the inequality constraints

$$\boldsymbol{A}\boldsymbol{x}^{\mathrm{T}} = \begin{bmatrix} 10 & -1 \\ -3 & 1 \\ -0.4 & 1 \end{bmatrix} \boldsymbol{x}^{\mathrm{T}} \leq \boldsymbol{b} = \begin{bmatrix} 9.1 \\ 0.1 \\ 0.7 \end{bmatrix} \qquad (25)$$

The inequality constraints are reformulated into equality constraints as

$$\boldsymbol{h}(\boldsymbol{x}) = \max(\boldsymbol{A}\boldsymbol{x}^{\mathrm{T}} - \boldsymbol{b}, \boldsymbol{0}) = \boldsymbol{0} \qquad (26)$$

In Figure 2, results from applying the algorithm to this problem are shown. In Figure 2a the initial nested design is shown. Black lines mark the constraint boundaries and the infeasible region is shown in grey. It is observed that five samples of the initial design violate constraints. Furthermore, the nested structure is apparent. The complete Latin hypercube design consists of 20 samples, which are marked with blue circles on the figure. A smaller Latin hypercube design of 10 samples, which is marked with red circles, is nested within this design. The sample design after annealing is shown in

Figure 2b. Notice that the constraints have now been resolved while the nested Latin hypercube structure has been preserved. The algorithm finished in approximately 172,000 iterations (for the gradient based optimizer a function evaluation is considered an iteration) taking 119 seconds in total on a laptop computer. While this may sound like a large number of iterations, it should be noted that the number of possible permutations of a design of this size is $(n_1! \cdot (n_2 - n_1)!)^k = (10! \cdot 10!)^2 \approx 1.73 \cdot 10^{26}$.

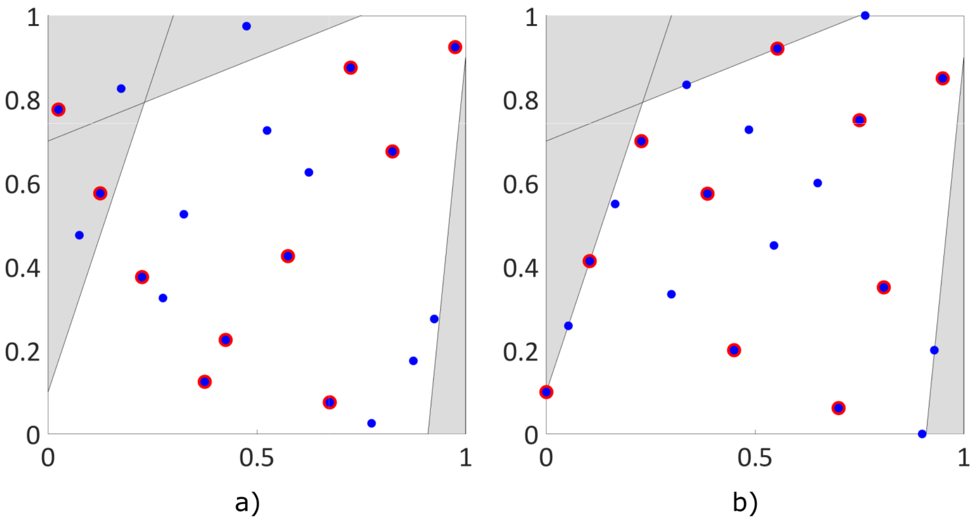

Figure 2. The proposed algorithm applied to a small two-dimensional problem. a) The initial sample design before annealing and b) the resulting sample design after annealing.

A small three-dimensional problem

A three-dimensional nested Latin hypercube with three layers of 5, 10, and 20 samples, respectively is produced on the unit cube. Inspired by the inequality constraints used above, a new set of three constraints in 3D space is defined as

$$\boldsymbol{A}\boldsymbol{x}^{\mathrm{T}} = \begin{bmatrix} 5 & 8 & -1 \\ -1.5 & -1.5 & 1 \\ -0.2 & 0.2 & 1 \end{bmatrix} \boldsymbol{x}^{\mathrm{T}} \leq \boldsymbol{b} = \begin{bmatrix} 12.1 \\ 0.1 \\ 1.03 \end{bmatrix} \tag{27}$$

In Figure 3 the nested sample design output by the developed algorithm is shown. Each constraint boundary surface is shown in the plot in translucent color, while the samples are color coded by layer in the following way: The first layer consists of the red samples, the second layer consists of the red *and* blue samples, and the third layer consists of *all* the shown samples, red, blue, and green. All three constraints have been satisfied by every sample in the design. The samples have been projected onto each axis in the plot. From these projections, it is apparent that each factor is sampled almost uniformly over the entire range. Furthermore, it should be noticed that each of the nested layers also has this property. It may be noticed that the uniformity is not so apparent in one of the axis in the figure. This is simply due to the inherent randomness in the design and can be improved upon by including more samples in the design or simply rerunning the algorithm. For this problem, the algorithm finished in approximately 277,000 iterations taking 187 seconds on a laptop computer.

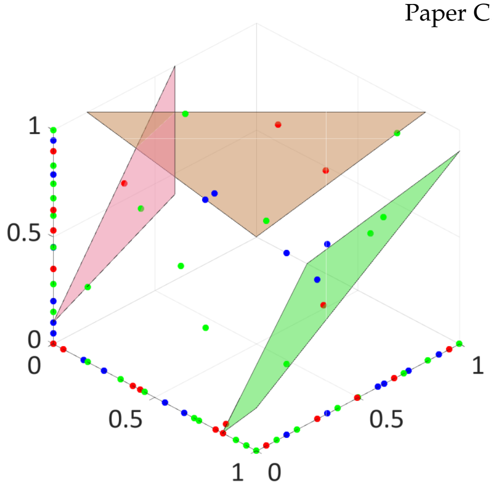

Figure 3. The developed algorithm demonstrated on a small three-dimensional example. The constraint boundary surfaces are shown as planes in translucent color. The individual samples are plotted both in the unit cube and as projected onto each axis. The samples of the first layer (5 samples) are shown in red. The second layer (10 samples) of the design consists of the samples plotted in blue *and* the samples of the first layer. The third and final layer (20 samples) consists of all the shown samples (red, blue *and* green).

An example application

Consider a variable shape mould of the type illustrated in Figure 4. A nested Latin hypercube sample design with two layers of 25 and 50 samples, respectively, in the space of geometries which are possible to produce on the mould is required. Since there are 25 actuators, this space is 25 dimensional. The number of possible permutations of this design is $(n_1! \cdot (n_2 - n_1)!)^k = (25! \cdot 25!)^{25} \approx 3.41 \cdot 10^{1259}$. The upper and lower bounds on the displacements of the 25 actuators are set to 0 mm and 100 mm, respectively, and the distance between actuators is 120 mm. Due to its mechanical properties, there is a limit to the allowed curvature of the interpolating membrane. This minimum allowed radius of curvature is set to $r_{min} = 400$ mm. To translate this into constraints, a set of natural cubic splines, interpolating over every column and row of actuators, is used to model the membrane. The curvature is evaluated at every actuator along both in-plane directions for a total of 50 curvatures. Since cubic splines and their derivatives are linear in the interpolated function values, the curvature constraints can be formulated linearly as

$$\boldsymbol{A}\boldsymbol{x}^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{A}_c \\ -\boldsymbol{A}_c \end{bmatrix} \boldsymbol{x}^{\mathrm{T}} \leq \begin{bmatrix} \boldsymbol{b}_c \\ \boldsymbol{b}_c \end{bmatrix} = \boldsymbol{b} \qquad (28)$$

where $\boldsymbol{A}_c\boldsymbol{x}^{\mathrm{T}}$ evaluates the curvatures at each actuator for the configuration $\boldsymbol{x}$, and $\boldsymbol{b}_c = 1/r_{min} * \mathbf{1}$ is a vector of the maximum allowed curvatures. The bottom half of the constraints is introduced since the constraints apply to both negative and positive curvatures. The total number of constraints for the sample design is therefore 5000 (50 samples with 100 constraints pr. sample). As above the inequality constraints are reformulated to equality constraints as

$$\boldsymbol{h}(\boldsymbol{x}) = \max(\boldsymbol{A}\boldsymbol{x}^{T} - \boldsymbol{b}, \mathbf{0}) = \mathbf{0} \qquad (29)$$

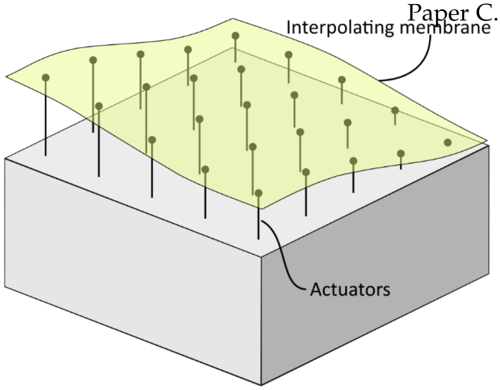

Figure 4. The principle in a variable shape mould. The mould consists of a bed of linear actuators and a flexible membrane which constitutes the mould surface.

For this problem, the algorithm finished in 4.2 million iterations taking 1 hour and 52 minutes on a laptop computer. In Figure 5 the evolution of the objective function and the sum of the constraint violations are plotted as a function of the iteration count. In Figure 5a it is observed that the constraint violation and the value of the Lagrangian follow each other approximately for most of the optimization, since the Lagrange multipliers were initialized as $\boldsymbol{\lambda} = \mathbf{1}$. When the number of violated constraints becomes small, spikes in the value of the Lagrangian begin to occur, because the values of some Lagrange multipliers are increased sharply to enforce satisfaction of the corresponding constraints. In Figure 5b a close-up of the final stage of the optimization is shown, where a gradient based optimizer takes over. It is apparent that some additional improvement of the design is possible by altering the location of each sample inside its bin (adjustment of $\boldsymbol{d\Pi}$, see Eq. (10)).

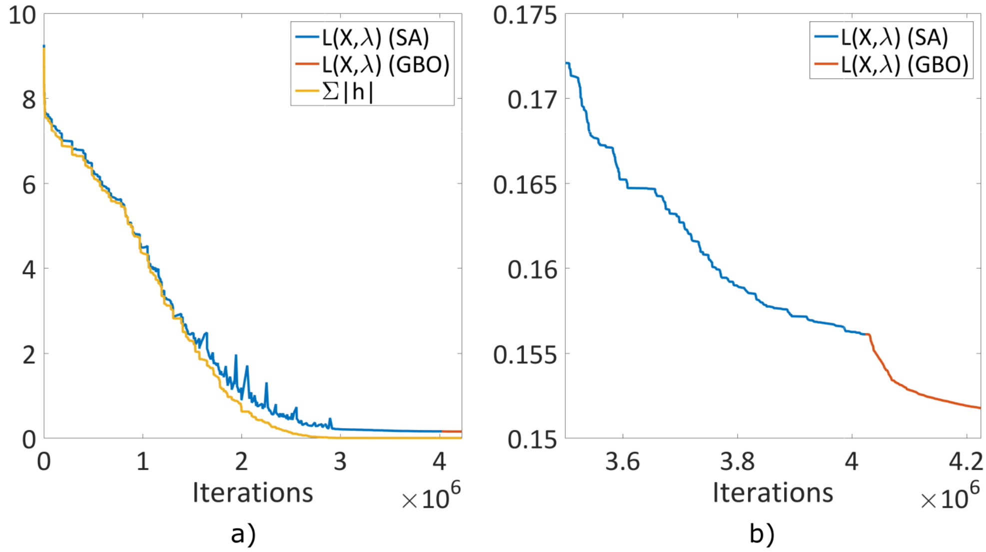

Figure 5. a) Performance of the proposed algorithm. b) Zoom on the final part where a gradient based optimizer takes over.

Conclusion

Latin hypercube sampling is a widely applied sampling strategy used across many disciplines and areas. In certain applications, such as multi-fidelity modeling, the specialized *nested* Latin hypercube sample designs are particularly useful. In the literature a lot of effort has been invested in the development of Latin hypercube sampling. However, in many cases, sampling cannot be performed freely over a complete hypercube, but must be constrained to an embedded subspace of the hypercube. Some literature concerning special forms of constraints has been found, but a method for generating Latin hypercube sample designs for more general sets of constraints has presented a knowledge gap.

In this paper this gap is sought closed by presenting an algorithm for producing high quality constrained orthogonal-maximin nested Latin hypercube sample designs.

The algorithm proceeds by solving a discrete Lagrange optimization problem using a simulated annealing algorithm. The objective function sought minimized is a Lagrangian function of the orthogonal-maximin Latin hypercube quality measure and a set of arbitrary constraints on the sample space.

The developed algorithm has been demonstrated on two examples and an application. It is shown that the algorithm is able to resolve violated constraints and improve the quality of an initial randomly generated nested Latin hypercube sample design, while persisting the nested Latin hypercube structure.

Nomenclature

v	A scalar.
$\boldsymbol{v}$	A column vector unless otherwise specified.
$\boldsymbol{v}_i$	The i'th entry of $\boldsymbol{v}$.
$\boldsymbol{V}$	A matrix.
$\boldsymbol{V}_i$	The i'th column/row of $\boldsymbol{V}$.
$\boldsymbol{V}_{i,j}$	Entry i,j of $\boldsymbol{V}$.
N	The total number of samples in the sample design.
n_i	The number of samples in layer i of the sample design.
t_i	The integer ratio $t_i = n_i/n_{i-1}$.
k	The number of factors in the sample design.

$\boldsymbol{X}$ A continuous sample design with scaled factors.

$\boldsymbol{d\Pi}$ Matrix specifying the continuous placements of samples within their respective bins in the sample design.

$\phi(\boldsymbol{X})$ The performance metric of the sample design $\boldsymbol{X}$.

w Parameter that determines the relative weighting of the correlation and maximin terms in ϕ.

p Parameter of ϕ.

$\boldsymbol{h}(\boldsymbol{X})$ A vector function of equality constraints.

$\boldsymbol{\lambda}$ A vector of Lagrange multipliers.

$\boldsymbol{d\lambda}$ Maximum rate of change of the Lagrange multipliers.

$L(\boldsymbol{X}, \boldsymbol{\lambda})$ The discrete Lagrangian function.

Z_q The set of integers from 1 to q.

$A \backslash B$ Operator which removes the intersection of A and B from A.

$G_X(\boldsymbol{X})$ Performs a perturbation of the sample design $\boldsymbol{X}$.

$G_\lambda(\boldsymbol{\lambda})$ Performs a perturbation of the Lagrange multipliers $\boldsymbol{\lambda}$.

T Temperature in the simulated annealing algorithm.

η Parameter of the simulated annealing algorithm.

τ Parameter of the simulated annealing algorithm.

β The probability that a Lagrange multiplier perturbation is performed in a given iteration.

n_T Number of iterations between evaluating if equilibrium at a temperature has occurred.

ϵ_{Accept} Parameters of the stopping criterion of the simulated annealing algorithm.

ϵ, N_ϵ

α Reduction factor for the cooling schedule of the simulated annealing algorithm.

References

Christensen, E.T., Forrester, A.I.J., Lund, E., and Lindgaard, E., 2018. Developing Meta-Models for Fast and Accurate Prediction of the Draping of Physical Surfaces. *Journal of Computing and Information Science in Engineering*, 18 (2).

Christensen, E.T., Lund, E., and Lindgaard, E., 2018. Experimental Validation of Surrogate Models for Predicting the Draping of Physical Interpolating Surfaces. *Journal of Mechanical Design*, 140 (1).

Forrester, A.I.J. and Keane, A.J., 2009. Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45 (1–3), 50–79.

Forrester, A.I.J., Sóbester, A., and Keane, A.J., 2008. *Engineering Design via Surrogate Modelling: A Practical Guide*. Wiley.

Joseph, V.R. and Ying, H., 2008. Orthogonal-Maximin Latin Hypercube Designs. *Statistica Sinica*, 18 (1), 171–186.

Lundy, M. and Mees, A., 1986. Convergence of an annealing algorithm. *Mathematical Programming*, 34 (1), 111–124.

McKay, M.D., Beckman, R.J., and Conover, W.J., 1979. Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, 21 (2), 239–245.

Morris, M.D. and Mitchell, T.J., 1995. Exploratory designs for computational experiments. *Journal of Statistical Planning and Inference*, 43 (3), 381–402.

Owen, A.B., 1994. Controlling Correlations in Latin Hypercube Samples. *Journal of the American Statistical Association*, 89 (428), 1517.

Petelet, M., Iooss, B., Asserin, O., and Loredo, A., 2010. Latin hypercube sampling with inequality constraints. *AStA Advances in Statistical Analysis*, 94 (4), 325–339.

Qian, P.Z.G., 2009. Nested Latin hypercube designs. *Biometrika*, 96 (4), 957–970.

Wah, B.W. and Wang, T., 1999. Constrained simulated annealing with applications in nonlinear continuous constrained global optimization. *In*: *Proceedings 11th International Conference on Tools with Artificial Intelligence*. IEEE Comput. Soc, 381–388.

ISSN (online): 2446-1636
ISBN (online): 978-87-7210-442-3
AALBORG UNIVERSITY PRESS